

# CHILDREN'S OPPORTUNITIES AND IMPACT EVALUATION

José Luis Figueroa Oropeza  
2015

Supervisor:  
Prof. Dr. Dirk Van de gaer

Submitted to the Faculty of Economics and Business Administration of Ghent University in  
fulfillment of the requirements of the degree Doctor in Economics



**Doctoral Committee:**

Prof. Dr. Marc De Clercq  
Dean-President, Ghent University

Prof. Dr. Patrick Van Kenhove  
Academic Secretary, Ghent University

Prof. Dr. Dirk Van de gaer  
Ghent University

Prof. Dr. Bart Cockx  
Ghent University

Dr. Francisco Ferreira  
The World Bank

Dr. Jef Leroy  
International Food Policy Research Institute

Dr. Bram Thuysbaert  
Bio-Invest/IPA



*A todos los redondos del mundo*



# Acknowledgements

“...If each of my words were a drop of water, you would see through them and glimpse what I feel: gratitude, acknowledgement” O.P.

Above all, my wholehearted gratitude to my Supervisor Dirk Van de gaer for leading me and being an excellent mentor during these 5 years. Dirk: thank you for endorsing the equality-of-opportunity principles and giving me so many chances to develop this PhD.

I also would like to thank the members of my reading committee for their time and valuable comments: Bart Cockx, Francisco Ferreira, Bram Thuysbaert, and Jef Leroy.

To my colleagues and friends at the Faculty of Economics: thanks for all the innumerable good moments.

Also, I would like to express my gratitude to my Belgian family for all their support since the first day I arrived to Belgium: Linda, Paul, Irmine, Joost, Lena and Sien, thank you all.

A mi familia: José Luis, Martha, Marco, César y Lilia: gracias, sin ustedes este largo camino se habría truncado hace mucho.

Finally, to Heleen and Félix, the two pillars of my entire life: thank you.

José Luis Figueroa Oropeza  
Ghent, December 2015





# Contents

<b>1</b>	<b>General Introduction</b>	<b>1</b>
1.1	Normative principles . . . . .	2
1.2	Empirical analysis of equality of opportunity . . . . .	3
1.3	Description of the evaluation sample . . . . .	6
1.4	Outline and main findings . . . . .	8
<b>2</b>	<b>Children's Health Opportunities and Project Evaluation</b>	<b>13</b>
2.1	Introduction . . . . .	14
2.2	Definitions and methodology . . . . .	17
2.3	Data description . . . . .	20
2.3.1	The Oportunidades program . . . . .	20
2.3.2	Sample Design . . . . .	21
2.3.3	Circumstances and outcomes . . . . .	23
2.4	Empirical Results . . . . .	26
2.4.1	Comparison of weighted treatment and control types . . . . .	26
2.4.2	Comparison to previous studies . . . . .	32
2.5	Conclusion . . . . .	34
	<b>Appendices</b>	<b>43</b>
2.A	Sampling procedure . . . . .	43
2.B	Results of the logistic regression . . . . .	43
2.C	Matching estimator and construction of the corresponding distribution function. . . . .	46
2.D	Testing stochastic dominance . . . . .	49
2.E	Roemer's identification axiom and matching estimator (weighted treatment distribution) . . . . .	50
2.F	Treatment and control effects in matched samples . . . . .	53
2.G	Sensitivity Analysis . . . . .	54
2.G.1	Entire delayed treatment group versus Control . . . . .	56
2.G.2	Mother's education as circumstance criterion . . . . .	62

2.G.3	Original versus Delay treatment . . . . .	62
<b>3</b>	<b>Distributional Effects of Oportunidades on Early Child Development</b>	<b>73</b>
3.1	Introduction . . . . .	75
3.2	Methods . . . . .	77
3.2.1	Data and Sample . . . . .	77
3.2.2	Propensity Score Matching . . . . .	78
3.2.3	Stochastic Dominance . . . . .	79
3.3	Outcomes . . . . .	81
3.4	Results . . . . .	82
3.5	Sensitivity Analysis . . . . .	85
3.6	Discussion . . . . .	88
	<b>Appendices</b>	<b>97</b>
3.A	Characteristics of the households in the sample in 1997 . . . . .	97
3.B	Matching procedure . . . . .	98
3.C	Sensitivity analysis . . . . .	101
<b>4</b>	<b>Did Progresa Reduce Inequality of Opportunity for School Re-enrollment?</b>	<b>107</b>
4.1	Introduction . . . . .	108
4.2	Description of Progresa and sample selection . . . . .	110
4.3	Methodology . . . . .	113
4.3.1	Theoretical framework . . . . .	113
4.3.2	Selection of circumstances . . . . .	116
4.3.3	Empirical methodology . . . . .	117
4.4	Empirical Results . . . . .	120
4.4.1	Average effect . . . . .	120
4.4.2	Dominance . . . . .	122
4.4.3	Decomposing the effect . . . . .	122
4.5	Conclusion . . . . .	125
	<b>Appendices</b>	<b>131</b>
4.A	Composition of the sample . . . . .	131
4.B	Bootstrap procedure . . . . .	148
4.C	Generalized Lorenz . . . . .	150
4.D	Lorenz . . . . .	151

# List of Figures

<b>2</b>	<b>Children’s Health Opportunities and Project Evaluation</b>	<b>13</b>
2.1	Stochastic dominance results. . . . .	30
2.C.1	Estimated propensity scores. . . . .	48
2.G.1.1	Estimated propensity scores (delay vs control). . . . .	60
2.G.1.2	Stochastic dominance results (delay vs control). . . . .	61
2.G.2.1	Estimated propensity scores (Mother’s education case). . . . .	65
2.G.2.2	Stochastic dominance results (Mother’s education case). . . . .	66
2.G.3.1	Estimated propensity scores (delay vs original treatment). . . . .	70
2.G.3.2	Stochastic dominance results (delay vs original treatment). . . . .	71
<b>3</b>	<b>Distributional Effects of Oportunidades on Early Child Development</b>	<b>73</b>
3.1	Test of first-order stochastic dominance . . . . .	81
3.2	Stochastic dominance results: Achen Index (Behavior problems) . . . . .	83
3.3	Stochastic dominance results: Woodcock-Johnson (short-term memory) . . . . .	84
3.4	Stochastic dominance results: Woodcock-Johnson (Visual integration) . . . . .	85
3.5	Stochastic dominance results: Communicative Development Inventories (CDI) . . . . .	85
3.6	Stochastic dominance results: Peabody Picture Vocabulary Test (PPVT) . . . . .	86
3.7	Estimated Propensity Scores by group . . . . .	100
3.8	Stochastic dominance results: Achen Index (Behavior problems) . . . . .	101
3.9	Stochastic dominance results: Woodcock-Johnson (short-term memory) . . . . .	101
3.10	Stochastic dominance results: Woodcock-Johnson (Visual integration) . . . . .	102
3.11	Stochastic dominance results: Communicative Development Inventories (CDI) . . . . .	102
3.12	Stochastic dominance results: Peabody Picture Vocabulary Test (PPVT) . . . . .	102
3.13	Stochastic dominance results: Achen Index (Behavior problems) . . . . .	103
3.14	Stochastic dominance results for children aged 2 years (Achen Index) . . . . .	104

**4 Did Progresá Reduce Inequality of Opportunity for School Re-enrollment?**107

4.1 Re-enrollment rates in October 1998 per grade for immediate and delayed treatment sample. . . . . 112

4.2 Generalized Lorenz Curve grade 6 . . . . . 123

4.3 Lorenz Curve grade 6 . . . . . 124

# List of Tables

<b>2</b>	<b>Children’s Health Opportunities and Project Evaluation</b>	<b>13</b>
2.1	Composition of the samples. . . . .	24
2.2	Health outcomes of 2-6 year old children in 2003. . . . .	25
2.3	Difference between control and treatment in fraction of anemic, stunted, at risk of being overweight and days sick. Weighted samples. . . . .	28
2.A.1	Sampling process. . . . .	43
2.B.1	Logistic regression results. . . . .	45
2.C.1	Propensity score matching: common support and number of observations in the common support. . . . .	46
2.F.1	Health outcomes of 2-6 year old children in 2003. . . . .	53
2.G.1.1	Composition of the samples (delay vs control). . . . .	56
2.G.1.2	Health outcomes of 2-6 year old children in 2003 (delay vs control). . . . .	57
2.G.1.3	Health outcomes of 2-6 year old children in 2003 (delay vs control): Matched samples. . . . .	57
2.G.1.4	Difference between control and treatment in fraction of anemic, stunted at risk of being overweight and days sick, weighted samples (delay vs control). . . . .	58
2.G.1.5	Logistic regression results (delay vs control). . . . .	59
2.G.1.6	Propensity score matching: common support and number of observations in the common support (delay vs control). . . . .	61
2.G.2.1	Composition of the samples (Mother’s education case). . . . .	62
2.G.2.2	Health outcomes of 2-6 year old children in 2003 (Mother’s education case). . .	62
2.G.2.3	Health outcomes of 2-6 year old children in 2003 (Mother’s education case): Matched samples. . . . .	63
2.G.2.4	Difference between control and treatment in fraction of anemic, stunted at risk of being overweight and days sick, weighted samples (Mother’s education case). .	63
2.G.2.5	Logistic regression results (Mother’s education case). . . . .	64

2.G.2.6	Propensity score matching: common support and number of observations in the common support (Mother's education case). . . . .	66
2.G.3.1	Composition of the samples (delay vs original treatment). . . . .	67
2.G.3.2	Health outcomes of 2-6 year old children in 2003 (delay vs original treatment). . . . .	67
2.G.3.3	Health outcomes of 2-6 year old children in 2003 (delay vs original treatment): Matched samples. . . . .	68
2.G.3.4	Difference between initial and delayed treatment in fraction of anemic, stunted at risk of being overweight and days sick, weighted samples (delay vs original treatment). . . . .	68
2.G.3.5	Logistic regression results (delay vs original treatment). . . . .	69
2.G.3.6	Propensity score matching: common support and number of observations in the common support (delay vs original treatment). . . . .	71
<b>3</b>	<b>Distributional Effects of Oportunidades on Early Child Development</b>	<b>73</b>
3.1	Observations before and after PSM. . . . .	79
3.2	Characteristics of the households in the sample in 1997. Logistic regression with dependent variable as 1 if observation belongs to treatment, 0 otherwise. . . . .	97
3.3	Propensity scores and number of observations by group . . . . .	99
3.4	Characteristics of the households in the matched sample. Logistic regression with dependent variable as 1 if observation belongs to treatment, 0 otherwise. . . . .	99
3.5	Average mean effect and descriptive statistics . . . . .	105
3.6	Quantile treatment effects at selected deciles <sup>(1)</sup> . . . . .	106

<b>4</b>	<b>Did Progresá Reduce Inequality of Opportunity for School Re-enrollment?</b>	<b>107</b>
4.1	Samples sizes and average (predicted) re-enrollment rates per grade in 1998 . . . . .	121
4.2	The Human Opportunity Index (all entries multiplied by 100) . . . . .	125
4.3	The Gini Social Welfare Function ( $S$ ) (all entries multiplied by 100) . . . . .	126
A.1	Composition of and missing values in treatment and control samples in percentage . . . . .	131
A.2	Logistic regression estimates . . . . .	132
A.3	Preprogram characteristics of the sample . . . . .	136
A.4	Preprogram characteristics of the sample . . . . .	139
A.5	Preprogram characteristics of the sample . . . . .	142
A.6	Preprogram characteristics of the sample . . . . .	145
A.7	Generalized Lorenz coordinates per grade for treatment and controls . . . . .	150
A.8	Lorenz coordinates per grade for treatment and controls . . . . .	151





# 1 | General Introduction

Program evaluation often adopts a positive rather than a normative perspective. “Impact” as understood in the evaluation literature, refers to the change of an outcome that can be attributed *exclusively* to the effect of the policy. The evaluation problem thus resides in extracting the causal effect of the program (Ravallion (2013)). However, such interpretation can be identified with a positive rather than a normative perspective, since the later, calls for value judgements of an ethical, political, or aesthetic nature (Ziliak, 2008). That is, an evaluation from a normative perspective looks at what should be instead of what it is. The purpose of my dissertation is to present an evaluation of a Conditional Cash Transfer program in Mexico from the perspective of children’s opportunities. Therefore, it provides an impact evaluation which is ultimately concerned with interpreting the results of the analysis on an ethical basis. To construct the bridge between purely impact evaluation and the broad concept of opportunity, I use methods from the literature on evaluation in a framework where ethical principles from the literature of equality of opportunity need to be satisfied.

Opportunity sensitive impact evaluations, above all, must consider justice the priority for the policy under scrutiny. Not long ago, the role of policies to enhance social justice was ignored, today however, most countries and multilateral organizations recognize the central role of public interventions to enhance opportunities for all human beings. For example, the United Nations Development Goals expressively incorporates access to opportunities in different domains as a universal goal (United Nations, 2008). Similarly, the World Development Report 2006 promotes the provision of opportunities for all, and the role of public policy as a key tool to “leveling the playing field” (The World Bank, 2005).

The development of the study of equality of opportunity began with contributions from political philosophers. John Rawls for example, associated the idea of justice with the equalization of opportunities in society (Rawls, 1971). The novelty of his proposal rested on the idea that opportunities, rather than outcomes, were the relevant space to judge social fairness. Further, his work opened up the modern study of equality of opportunity and was the foundation for many influential works that came later. Other contributions that paved the way on equality of opportunity soon followed that of Rawls. Sen (1980) for example, argued

that social policy should promote equality of *capabilities*. For Sen, individual achievement should not rely on people's capacities but only on the choices made by individuals. Sen refers to the set of such capacities as "functionings", which among others, include education and health. A bit later, [Dworkin \(1981a,b\)](#) claimed for equalization of resources, and according to him, the role of social policies rests on the equalization of resources in society. Finally, two influential works by [Cohen \(1989\)](#) and [Arneson \(1989\)](#) defend the idea of placing choice and responsibility at the center of the evaluation of equal-opportunity policies.

Among economists, the incorporation of normative principles to the analysis of policies, particularly from an equal-opportunity perspective is more recent. Such interest, has mainly led to a large number of contributions on methods and approaches to measure inequality of opportunity. However, the diversity of approaches often complicates the empirical analysis, and only recently, proposals to unify the literature on methodological aspects and their relation with normative principles start to appear in the literature. For example, [Ramos and Van de gaer \(2015\)](#) provides a thoroughly overview on normative principles on the one hand, and measures on the other. Further, they discuss how principles accommodate in empirical analysis available in the literature. Another contribution which organizes the existent literature according to equal-opportunity principles is provided by [Ferreira H.G. and Peragine \(2015\)](#). They provide a "canonical" model which embodies the main concepts and methodological theoretical aspects. Finally, Roemer and Trannoy's Chapter in the Handbook of Income Distributions extensively discusses philosophical, methodological, and empirical aspects of the theory of equality opportunity ([Roemer and Trannoy, 2014](#)).

In what follows I discuss the main principles found in the literature of equality of opportunity and at the same time, I explain how these principles are framed and used in each the three Chapters presented in the dissertation.

## 1.1 Normative principles

To incorporate opportunity principles into impact evaluation analysis one needs, first of all, to consider the process that generates individual outcomes. Impact evaluation methods rely on comparisons between outcomes in different states (usually between treated and non-treated states), without placing too much emphasis on the role that individuals play in generating these outcomes. That is, the main point of impact evaluation analysis is the effect of the policy or program under study, and less how individuals with different characteristics react to the policy. Equal-opportunity analysis, in contrast, requires to discriminate between effects that are morally acceptable, because individuals are somehow responsible for them, and effects that

are not. The reason is that at the core of the equality of opportunity principle rest the idea that some inequalities are not offensive because individuals are up to some degree responsible, and as such, not all inequalities need to be eliminated. Two fundamental principles postulated in the literature of equality of opportunity formally recover the latter idea as follows:

The *reward principle* states that effort should be adequately rewarded. If individuals exert effort according to their own preferences, from an ethical viewpoint, they are entitled to enjoy the by-product of such effort. As a consequence, inequalities resulting from some individuals being more diligent than others should be respected. To invoke the *reward principle* two assumptions are due: first, individuals freely and consciously choose to exert their preferred level of effort; and second, individual effort is independent of characteristics beyond personal control<sup>1</sup>. As an example of the former, consider the argument that beyond disadvantageous conditions in life, individuals choose how much time and effort to spend studying or doing homework (providing of course that they had the opportunity to attend school). The second principle on the other hand, the *compensation principle*, says that inequalities arising from differences in circumstances (i.e., non-responsibility factors) are unjust, and therefore, they need to be eliminated. Compensation is due in situations where equally diligent individuals do not achieve the same outcomes. Both, reward and compensation principles, rest on the idea that responsibility is important when judging differences between individuals, which finally implies that equality of opportunity arises when inequalities among individuals are only due to differences in individual effort.

## 1.2 Empirical analysis of equality of opportunity

Operating these principles empirically requires to deal with a number of issues. As exposed above, equality of opportunity entails ignoring morally irrelevant sources of inequalities to focus exclusively on those beyond personal control. In empirical analyses, this calls for a clear classification between non-responsibility factors, referred in the literature as *circumstances*, and responsibility factors, or *efforts*. Most characteristics that can be considered as circumstances refer to parental background, and as such, are incorporated in the different chapters of the dissertation. Information on parental education, for example, is taking into account in the three analyses presented below because of the importance for children's future well-being (see, e.g., Bourguignon, Ferreira, and Menéndez (2007)). Similarly, other characteristics, like race or gender, have been consistently used in the literature as non-responsibility factors,

---

<sup>1</sup>The majority view accepts independence of effort and non-responsibility characteristics, however, responsibility for preferences might lead to different conclusions if preferences are not fully endorsed by individuals, in which case one should take some of these preferences as circumstances.

and as such, are also part of the set of circumstances used in the thesis. In the case of gender, the evidence suggests that in Mexico, girls and indigenous are relegated in terms of education, labor, and in general, in terms of development opportunities (Busso, Cicowiez, and Gasparini (2005), Lewis and Lockheed (2008)). Other characteristics, such as place of residence and availability of services in the locality are also considered and discussed in turn in each chapter.

Less clear however, is how to classify and identify factors that involve some degree of personal control since one could argue that some preferences are partly determined by circumstances. For example, some might consider smoking as a circumstance if growing in less educated families is associated with smoking preferences, and therefore, they might support the idea of compensating inequalities resulting from such unhealthy preferences. Others, in contrast, might consider this unacceptable and justify disadvantages that smokers could suffer. In the analysis presented here the only possible responsibility attached to children is the responsibility of their parents. In Chapter 1, this point is particularly important because unobserved factors affecting children's health, like genetic luck, are treated as compensation factors depending on interpretations arising with dominance criteria. For instance, first-order dominance in favor of the program implies that the role of parental responsibility is respected (only compensation due to circumstances are relevant), whereas second-order can be obtained by within-type transfers which implies not respecting parental effort. Finally, an assumption made in the analysis is that program participation is beyond children's control, and as such, is considered a circumstance. This assumption is crucial because from the perspective of children's opportunities, the program may compensate or not participant children, and from the evaluation viewpoint is important that participation is completely exogenous. As explained below, program participation was voluntary, but the program was offered to families, not to individuals. It results very unlikely that children living in these households could have decided by themselves to participate, therefore, one can plausibly assume that participation is indeed beyond children's control.

Another issue in empirical applications refers to data constraints since a complete list of circumstances is rarely available, whereas efforts, are usually not observable. The former is problematic when the functional relation between outcomes and circumstances is modeled through a parametric approach. Ferreira and Gignoux (2011) find that the effect of taking into account a subset of circumstances leads to a lower-bound estimation of the true level of inequality of opportunity. Niehues and Peichl (2014) find similar results by computing upper-bound estimates. The conclusions from both studies suggest the effect of unobserved circumstances, such as ability or talent, account for a large part of the inequality of oppor-

tunity. In the literature, solutions to overcome the lack of information on individual effort have been suggested. For instance, it is possible to adopt an ex-ante approach and only look at individual prospects, since an ex-ante perspective only takes into account the effect of circumstances. Alternatively, one could follow an ex-post approach and rely on Roemer's Identification Axiom (Roemer, 1993) to infer individual effort from distributions of types. I elaborate on this point next.

Broadly, taking an ex-ante or an ex-post approach depends on the possibility of extracting information about individual effort. An ex-ante perspective is compatible with the compensation principle exposed above, because it consists in observing individuals *before* they exert effort. From an ex-ante point of view, it is sufficient to observe individual circumstances, and for some authors, only ex-ante inequalities should be compensated by public intervention (see, e.g., Pignataro (2012)). Because of that, much of the contributions in the literature on measurement of inequality of opportunity have adopted an ex-ante rather than an ex-post approach. In Chapter 3 for example, the analysis departs from an ex-ante perspective since only children's circumstances are taken into account.

Carrying out an ex-post analysis in contrast, demands observing information about individual effort. Because effort is hard to observe, doing ex-post analysis requires solutions to infer it. An elegant solution by John Roemer, the so-called Roemer's Identification Axiom (RIA), proposes to recover information on individual responsibility by comparing individuals sharing the same set of circumstances (i.e., have the same type). Under RIA, two individuals are equally responsible if they are located at the same quantile of their type's distribution. In other words, if two individuals share the same non-responsibility characteristics (circumstances), and one of them is located in a better position in the distribution of, say income, that comes only by differences in effort. The first two chapters depart from an ex-post perspective, but RIA is only explicitly applied in Chapter 1. Importantly, for the analysis in Chapter 1 as explained before, parental effort is the normatively relevant responsibility characteristic since children, by definition, are not accountable for their actions yet.

Last point, but not less relevant, is the methodological strategy. In the literature, two approaches dominate. On the one hand, one can model parametrically the relation of outcomes and individual characteristics by assuming a functional form. This approach has the advantage of allowing the use of multiple variables in the right hand side of the equation which means that more types can be used in the analysis. This is also the approach followed in Chapter 3. However, using a functional form requires doing assumptions about the relationship between outcomes and circumstances, which could mislead the true nature of the relationship. In Chapter 3 a logistic model is used to estimate the probability of school re-

enrollment, but other specifications might be adequate as well.<sup>2</sup> In contrast, a non-parametric approach does not require to make assumptions about functional relations, but its implementation requires a sufficiently large sample size. Also, to carry out non-parametric methods, like stochastic dominance as presented in Chapters 1 and 2, one needs to make pairwise comparisons between types, which restrict the number of types that can be taken into account in the analysis. For example, to construct types based on gender, race, and parental education (assuming education a dummy) one obtains 8 possible type-combinations. Adding one extra dummy characteristic like access to school, translates into 16 types, and so forth. Therefore, this kind of analysis does not provide a comprehensive assessment of the effect of circumstances on opportunities but only which types gain on a one-to-one basis. In that sense, the analysis presented in Chapter 3 results much more complete as it takes on board a large list of individual characteristics.

### 1.3 Description of the evaluation sample

Oportunidades, formerly named Progresa, and only recently re-named as Prospera in 2014, is one of the largest Conditional Cash Transfer programs in the world. Originally implemented in rural communities, the program covers more than 5.8 million families in Mexico. As the backbone of the anti-poverty agenda of the Mexican government, the program aims at breaking the intergenerational cycle of poverty by promoting human capital accumulation, especially for children, and by providing in-kind and monetary transfers to alleviate current poverty. Investments by the program are mainly done in three areas: health, education, and nutrition of all family members, but special attention is given to infants, lactating and pregnant women. Additionally, the program grants scholarships for each child attending school, which represents around one fifth of consumption of an average poor family in rural areas in Mexico (Levy, 2007)<sup>3</sup>.

Since the onset of the program, counting with a rigorous evaluation has been a priority for the Mexican government. Before the implementation of Oportunidades, very few evidence about the effect of public programs to eradicate poverty was available. Stemmed from the need to generate reliable evidence for policy makers, the original design of the program contemplated a comprehensive evaluation based on an experimental design. During the experimental phase of the program, information on families which were randomly selected for participation

---

<sup>2</sup>Although not reported in the text, I performed a test to check the linearity of the index function by including higher-order terms in the specification of the logit model in Chapter 3. The null that the model is correctly specified was not rejected at conventional levels of statistical significance. See Wooldridge (2010), pp. 570.

<sup>3</sup>This information refers only to the average rural household in 1998

was collected. The analysis of Chapter 3 refers to a group of children that belongs to the original evaluation sample collected during the initial phase. The experimental nature of the evaluation consisted in randomly *offering* the program to families; therefore, only families that decided to participate are observed in the data. This also implies that comparisons between treated and non-treated families provide an intend-to-treatment analysis.<sup>4</sup> Collection of data during the experimental phase followed the next steps:

First, 506 localities with high concentration of poor households were selected to be included in the program. All selected localities were situated in rural areas and had less than 2,500 and more than 500 inhabitants. Then, 320 out the 506 localities were randomly assigned to treatment and the remaining 186 to control. Within localities, households that according to criteria from the Mexican Statistics Office (Instituto Nacional de Estadística y Geografía, INEGI) were classified as poor, were eligible to enter the program and eventually were offered to participate (INSP, 2005). Eligible households from control localities, on the other hand, were left out of the program and served as comparison group. These two groups comprise the original experimental sample, and are referred to in the different chapters as the “immediate” and “delayed” treatment groups respectively.

In contrast, the samples of Chapter 1 and 2 come from the non-experimental phase of the program that came after the original control (delayed treatment) was incorporated. After two years of operations, in the year 2000, the original control group (delayed treatment) was incorporated to the program, and with this incorporation, the original experimental design was lost.<sup>5</sup> Therefore, for subsequent evaluations it was necessary to have a new control group. The new control group was constructed by a Propensity Score Matching procedure at locality level, and it resulted in the selection of 151 new localities that were similar to the original 506 original localities of the experimental phase (INSP, 2005). The new control however, only served as the basis for selecting the comparison groups in Chapter 1 and 2. There were a number of reasons for that. First, the analysis presented in these chapters requires to construct types of children (groups of children sharing the same non-responsibility characteristics) and the Propensity Score Matching to construct the control was done at locality instead of individual level. Second, the Propensity Score was based on characteristics in 2000, when the delayed treatment group was already enrolled to the program. Because the Propensity Score should reflect the propensity of participation based on characteristics that are not influenced by program, a “second-layer” Propensity Score Matching was implemented

---

<sup>4</sup>However, as documented by Gertler (2004), program take-up was practically universal: as much as 97% of the families who were eligible decided to participate. Therefore, intend-to-treatment and treatment-on-the-treated analysis are very close in this case.

<sup>5</sup>Political reasons forced the Government to offer the program in the original control communities sooner that it was originally planned

at individual level between types. In Chapter 1 this translates into 4 matchings between treated and non-treated types (IL, IP, NL, NP types), and in Chapter 2 into 5 matchings (Complete sample, girls, boys, indigenous, and non-indigenous).<sup>6</sup> Finally, only children from households, that according to administrative records, received monetary transfers at the time of the evaluation were considered for the analyses of Chapters 1 and 2. Such restriction avoids the inclusion of children from the treatment group that in reality, were never treated, and thus it leads to a treatment-on-the-treated analysis.

## 1.4 Outline and main findings

The 3 chapters presented here revolve around the same question, namely, what is the effect of the Oportunidades program on children's opportunities in different domains and at the same time, which methods should be used in this context. I focus on children's outcomes because this group is of special interest for the program, and because opportunities early in life (or the lack of them) determine what children can achieve, be, or enjoy in adulthood.

Chapter 1 presents the analysis of health outcomes for children aged 2-6 years. To establish the effect of the program, conditional cumulative distribution functions (CDF's) for children in and out the program are compared according to stochastic dominance criteria. As argued in this Chapter, from an opportunity perspective, comparisons between CDF's should be based on first and second-order dominance, which requires invoking Roemer's Identification Axiom explained above. This, as an alternative to Perfect Positive Quantile Dependence (Heckman, Jeffrey, and Clements, 1997) which is normally required in impact evaluation analysis using CDF's. Also, to operationalize RIA in this context, it is necessary to compare children whose parents share the same non-responsibility characteristics (given that children are not responsible individuals).

On the other hand, the use of stochastic dominance methods, allows identifying where in the distribution the program is effective for children whose parents have certain characteristics or types, as is referred to in the literature of equality of opportunity. The proposed methodology thus contributes to both, the literature on equality of opportunity and the literature on impact evaluation. The results on this Chapter suggest that the treatment has substantial positive effects on the health opportunities for children from indigenous parents. Also, some effects are found on non-indigenous children, although these are weaker.

Chapter 2 offers an evaluation on cognitive and non-cognitive early child development. The focus of this chapter is to emphasize the distributional effects of the program. Although

---

<sup>6</sup>preprogram information for the control group came from recollection questions in 2003 about 1997. For more details refer to Section 2 in Chapter 1 and Section 2.2 if Chapter 2



less attention is paid here to normative principles, the methodology, as in the previous chapter is inspired in normative principles of equality of opportunity. In contrast with Chapter 1, the analysis does not rely on the use of types, but rather on comparisons between children with different characteristics. The relatively small number of observations in the sample impedes the construction of types (remember that sufficiently large number of observations are needed for non-parametric methods).

Four indicators of cognitive and one of non-cognitive ability are analyzed. The results suggest the program positively influences children's non-cognitive abilities: children enrolled in the program manifest fewer behavioral problems compared to those not enrolled. Effects for boys and girls, and for indigenous and non-indigenous children are substantial and cover a large part of the outcome's distribution. With regard to cognitive development, results are less outspoken: only short-term memory ability shows positive effects. Nonetheless, the results demonstrate that children with low values of cognitive development benefit from the program, whereas children with high values do not. From an equal opportunity perspective, these results can be interpreted as positive.

Chapter 3 presents the analysis of opportunities for school re-enrollment for children aged 6-16. The main contribution of this chapter is to offer a systematic, more comprehensive, assessment of the effect of the program on children's opportunities. Opposed to the nonparametric methods of Chapter 1 and 2 which do not allow the inclusion of a large number of circumstances, the parametric approach followed here permits including a large set of circumstances which brings a much more complete assessment for different types. Another point of contrast with the previous two chapters is the use of counterfactual distributions. The peculiarity in the use of counterfactuals relies on the fact that only differences due to circumstances are reflected, and therefore, only differences due to unequal opportunities are contained in these distributions. Further, to evaluate the effect of the program, two counterfactuals are constructed: one in case children are treated, and one in case they are not. To construct the counterfactuals, it is necessary to obtain a measure that reflects the level of opportunities faced by the children. In this case, the measure is the probability of school re-enrollment, which, as explained before, is calculated using a logistic model. A main assumption of the Chapter, as exposed in the methodological section, is that these probabilities give the values of children's opportunities for school re-enrollment.

The evaluation of the effect of the program on children's opportunities in Chapter 3 is done in two ways. First, I use functions to map the probability of school re-enrollment of each child into a real number, and impose desirable properties to this function. These properties (anonymity, non-decreasingness, the weak transfer principle and relativity) are

standard in the axiomatic literature of equality of opportunity and are discussed in detailed in the methodology section of the chapter. Providing that these properties are fulfilled, higher aggregate opportunities will arise due to the program if the distribution of treated children Generalized Lorenz dominates the one of non-treated, and similarly, one can claim that there is less inequality due to the program if the distribution of treated Lorenz dominates that of non-treated. Second, in order to decompose the improvement and the inequality effects of the Generalized Lorenz and Lorenz dominance results, two abbreviated opportunity functions are proposed: Human Opportunity Index (de Barros, Vega, and Saavedra, 2008) and the Gini Index. Both show how much of the effect is due to an average gain (level effect) and how much is due to a redistribution effect (Lambert, 2001).

The results of Chapter 3 are the following: 1) Progresa increases aggregate school re-enrollment opportunities since the distribution conditional on circumstances under treatment Generalized Lorenz dominates that when non-treated. 2) Progresa decreases inequality of re-enrollment opportunities, since the distribution in case of treatment, Lorenz dominates the distribution in case of not being treated. This results holds for any measure of relative inequality. And 3), the effect of the reduction of inequality of opportunity due to program participation is between 15 and 40% when looking at the Human Opportunity Index or the Gini Index. Finally, the results suggest children in the transition between primary and secondary education are particularly prone to benefit from the program. Also, for these children, the decrease in inequality of opportunity is markedly large. Given that many children abandon school after completing primary education in Mexico, interventions like Oportunidades not only decrease average dropout but have strong distributive effects for those most disadvantaged. However, these results apply only to the 25% of the population in Mexico incorporated to the program, which limits the reach of the conclusions for the entire population. In that sense, future studies that incorporate, for example, general equilibrium effects could inform about the effect of Conditional Cash Transfer programs on inequality of opportunity for the entire population.

## Bibliography

- ARNESON, R. J. (1989): "Equality and equal opportunity for welfare," *Philosophical studies*, 56(1), 77–93.
- BOURGUIGNON, F., F. FERREIRA, AND M. MENÉNDEZ (2007): "Inequality of opportunity in Brazil," *Review of Income and Wealth*, 53(4), 585–618.
- BUSO, M., M. CICOWIEZ, AND L. GASPARINI (2005): "Ethnicity and the Millennium Development Goals in Latin America and the Caribbean," *Documentos de Trabajo del CEDLAS*.
- COHEN, G. A. (1989): "On the currency of egalitarian justice," *Ethics*, pp. 906–944.
- DE BARROS, R. P., J. R. M. VEGA, AND J. SAAVEDRA (2008): "Measuring inequality of opportunities for children. Unpublished.," *The World Bank*.
- DWORKIN, R. (1981a): "What is equality? Part 1: Equality of welfare," *Philosophy & public affairs*, pp. 185–246.
- (1981b): "What is equality? Part 2: Equality of resources," *Philosophy & Public Affairs*, pp. 283–345.
- FERREIRA, F. H. G., AND J. GIGNOUX (2011): "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America," *Review of Income and Wealth*, forthcoming.
- FERREIRA H.G., F., AND V. PERAGINE (2015): "Equality of opportunity: Theory and evidence," *Policy Research Working paper, No 7217. The World Bank. Washington, D.C.*
- GERTLER, P. (2004): "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment," *American Economic Review*, 94(2), 336–341.
- HECKMAN, J., S. JEFFREY, AND N. CLEMENTS (1997): "Making the Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64(4), 487–535.
- INSP (2005): "General Rural Methodology Note," *Instituto Nacional de Salud Pública*, INSP2005.
- LAMBERT, P. J. (2001): *The Distribution and Redistribution of Income: Third edition*. Manchester University Press.

- LEVY, S. (2007): *Progress against poverty: sustaining Mexico's Progresa-Oportunidades program*. Brookings Institution Press.
- LEWIS, M., AND M. LOCKHEED (2008): "Social exclusion and the gender gap in education," *World Bank Policy Research Working Paper Series*, Vol.
- NIEHUES, J., AND A. PEICHL (2014): "Upper bounds of inequality of opportunity: theory and evidence for Germany and the US," *Social Choice and Welfare*, 43(1), 73–99.
- PIGNATARO, G. (2012): "Equality of opportunity: Policy and measurement paradigms," *Journal of Economic Surveys*, 26(5), 800–834.
- RAMOS, X., AND D. VAN DE GAER (2015): "Approaches to inequality of opportunity: Principles, measures, and evidence," *Journal of Economic Surveys*.
- RAVALLION, M. (2013): "The Idea of Antipoverty Policy," Discussion paper, National Bureau of Economic Research.
- RAWLS, J. (1971): *A Theory of Justice*. Harvard University Press, Cambridge, MA.
- ROEMER, J. (1993): "A Pragmatic Theory of Responsibility for the Egalitarian Planner," *Philosophy & Public Affairs*, 22(2), 146–166.
- ROEMER, J. E., AND A. TRANNOY (2014): *Equality of Opportunity* chap. 4, pp. 217–300. North-Holland.
- SEN, A. (1980): *Equality of what?*, vol. 1 of *The Tanner Lecture on Human Values*. Cambridge University Press, Cambridge.
- THE WORLD BANK (2005): "World Development Report 2006: Equity and Development," Discussion paper, The World Bank, Washington, D.C.
- UNITED NATIONS (2008): *The Millennium Development Goals Report 2008*. United Nations Publications.
- WOOLDRIDGE, J. M. (2010): *Econometric analysis of cross section and panel data*. MIT press.
- ZILIAK, S. T. (2008): *Normative Social Science and Novelty in Economics* chap. 1, pp. 535–536. Thomson Gale.

## 2 | Children's Health Opportunities and Project Evaluation: Mexico's Oportunidades Program

with Dirk Van de gaer and Joost Vandenbossche

Published in *The World Bank Economic Review*. Vol. 28, N.2, pp.282-310.

**Abstract:** We propose a methodology to evaluate social projects from the perspective of children's opportunities on the basis of the effects of these projects on the distribution of outcomes. We condition our evaluation on characteristics for which individuals are not responsible; in this case, we use parental education level and indigenous background. The methodology is applied to evaluate the effects on children's health opportunities of Mexico's Oportunidades program, one of the largest conditional cash transfer programs for poor households in the world. The evidence from this program shows that gains in health opportunities for children from indigenous backgrounds are substantial and are situated in crucial parts of the distribution, whereas gains for children from nonindigenous backgrounds are more limited.

## 2.1 Introduction

This paper evaluates the change in health opportunities for children aged two to six years who participate in the Mexican Oportunidades program. Oportunidades is a large-scale, conditional cash transfer program initiated in 1998 through which poor rural households receive cash in exchange for their compliance with preventive health care requirements, nutrition supplementation, education, and monitoring. In 2010, approximately 5.8 million families participated in the program, and cash transfers to the participants totaled \$4.8 billion. The average treatment effects of the program on the health of young children have been shown to be positive (see the literature surveyed in [Parker, Rubalcava, and Teruel \(2008\)](#)). We propose a methodology that focuses on the conditional cumulative distribution functions of health outcomes to identify whether and where in the distribution the program is effective for children whose parents have certain characteristics. Our methodology evaluates the program from the perspective of children's opportunities rather than average treatment effects.

[Fiszbein, Schady, and Ferreira \(2009\)](#) report that in 1997, only three developing countries (Mexico, Brazil, and Bangladesh) had conditional cash transfer programs in place; by 2008, this number had increased to 29, with many more countries planning to implement such programs. It is important to develop techniques to evaluate the effects of these programs on children's opportunities, because these programs are increasingly popular in developing countries, they are sometimes conducted on a large scale, and their focus is on breaking the intergenerational poverty cycle. Despite the recent emergence of substantial empirical literature measuring inequality of opportunity (e.g., 2009 [de Barros, Ferreira, Molinas Vega, and Chanduvi \(2009\)](#) and the references below), no such techniques currently exist.

In the recent literature on equality of opportunity (e.g., [Bossert \(1995\); Fleurbaey \(1995\), Fleurbaey \(2008\); Roemer \(1993\)](#)), a distinction is generally drawn between two types of factors that influence the outcome under consideration. On the one hand, there are circumstances, characteristics for which an individual is not responsible, such as race, sex, and parental background; these are the characteristics upon which we condition the cumulative distribution function. On the other hand, there are other characteristics for which individuals are considered responsible, such as having a good work ethic. The idea is that public policies, including conditional cash transfer programs, should compensate for the former while respecting the influence of the latter.<sup>1</sup>

We apply the framework to health outcomes of children aged two to six years. We con-

---

<sup>1</sup>Recently [Lefranc, Pistoiesi, and Trannoy \(2009\)](#) extend this framework with a third factor, random factors that are legitimate sources of inequality "as long as they affect individual outcomes and circumstances in a neutral way" (p. 1192)

sider the following circumstances for which parents are not responsible: race, in particular, whether either parent is indigenous; educational level, determined by whether either parent had primary education; and participation in the program. Each possible combination of circumstances corresponds to a “type,” in Roemer’s terminology (Roemer, 1993). Therefore, we have eight types. To evaluate the program, we take the health outcomes of children who belong to families enrolled in the program for each of the four types, which are defined on the basis of the parents’ race and education level, and we compare those outcomes with the health outcomes of children whose parents belong to the corresponding type that was not enrolled in the program. Within each type, outcomes can (and will) differ because of factors that are unobserved and ascribed to parental responsibility, such as parental health investments in children. In section 2.2, we argue that an opportunity perspective implies that the comparison of treatment and control types must be based on first- or second-order stochastic dominance.

The idea of using first- or second-order stochastic dominance to investigate equality of opportunity for a particular outcome is not novel. However, until now, this method has been applied only to study whether opportunities are equal within a particular population (see O’Neill, Sweetman, and Van de gaer (2000) and Lefranc, Pistolesi, and Trannoy (2009) for studies in which the outcome is income; see Rosa Dias (2009) and Trannoy, Sandy, Jusot, and Devaux (2010) for adults’ self-assessed health studies; for comparisons between different countries, see Lefranc, Pistolesi, and Trannoy (2008) for income-based outcomes; for comparisons between regions, see Peragine and Serlenga (2008) for education-based outcomes). Our paper makes three primary contributions to this literature. First, and most important, we conduct our evaluation by establishing the effect of Oportunidades on children’s health opportunities. Second, we consider opportunity in the health of young children because their health is crucial for their adult outcomes (see, e.g., Black, Devereux, and Salvanes (2007) and Alderman, Hoddinott, and Kinsey (2006)) and because it is important in its own right. Third, in contrast to previous literature that tested for stochastic dominance in the context of equality of opportunity, our test procedure is based on Davidson and Duclos (2009) and Davidson (2009). Thus, we test the null of nondominance against the alternative of dominance so that rejection of the null logically entails dominance.

Most of the literature on program evaluation focuses on estimating average treatment effects. However, we are interested in establishing or rejecting stochastic dominance between the distributions of health outcomes of children when their parents are either in or out of the program. This exercise is not trivial because we cannot observe the same child both in and out of the program; in other words, we cannot simply resort to a comparison of the cumulative distributions of treatment and control types without making additional assumptions

(Heckman, 1992). One such assumption is perfect positive quantile dependence (see Heckman, Jeffrey, and Clements (1997)), which stipulates that those who are at the  $q$ th quantile in the distribution with treatment would have been at the  $q$ th quantile in the distribution without treatment. Roemer's identification axiom (Roemer, 1993) is usually invoked in empirical applications of equality of opportunity when responsibility characteristics are unobserved. This axiom posits that the parents of children who are at the same percentile of their type distribution have exercised comparable responsibility. We argue below that this axiom provides a normatively inspired alternative to perfect positive quantile dependence by reducing the problem to a comparison of the cumulative distribution functions of the corresponding treatment and control types. The literature on average treatment effects stresses that treatment and control samples must be comparable in terms of preprogram characteristics. We show that this is also imperative when testing for stochastic dominance. Following the literature on average treatment effects, we propose a propensity score matching technique on the basis of preprogram characteristics to better compare treatment and control types. Finally, it is noteworthy that two authors recently suggested incorporating stochastic dominance into project evaluation: Verme (2010) proposed a stochastic dominance approach to determine the effect of a perfectly randomized experiment based on the measures establishing poverty line dominance (i.e., dominance for a range of poverty lines) developed by Foster, Greer, and Thorbecke (1984). Our approach, based on equality of opportunity, stresses that we should focus on the distributions that are conditional on circumstances instead of comparing the distributions of all treatment and control samples. Therefore, we compare the distributions of corresponding treatment and control types. Moreover, our propensity score matching technique makes this approach effective for imperfectly randomized experiments. Naschold and Barrett (2010) allow for nonrandomized treatment by focusing on stochastic dominance between treatment and control samples of the distribution of the difference in outcome, both before and after treatment. They do not focus on types, and the results are difficult to interpret because dominance in terms of differences does not imply that treatment leads to a dominating distribution, which fundamentally depends on who gains and who loses.

Our main findings are that the treatment has substantial positive effects on the health opportunities of children from indigenous families. The effects on children growing up in nonindigenous families are weaker, although we still find significant positive treatment effects for that group.

The paper is structured as follows. Section 2.2 provides definitions and explains the methodology. The data are described in section 2.3. Section 2.4 presents the empirical results, including a discussion of the relationship with previous studies. Section 2.5 concludes.



## 2.2 Definitions and methodology

Let a child's health outcome be represented by the variable  $h \in H = [\underline{h}, \bar{h}] \subseteq \mathbb{R}$  and let higher values of  $h$  mean better health. A child's health is the result of two types of variables. The first variable,  $c \in C$ , represents circumstances and characteristics for which the child's parents are not responsible, such as race, educational background, and whether the family participates in the program<sup>2</sup>) The second variable,  $r \in R$ , represents characteristics for which parents are responsible, such as health investments in children. Each combination of circumstances corresponds to a type. Social programs should improve children's opportunities, and from the perspective of the equality of opportunity literature, they should compensate for health differences that are caused by circumstances. Moreover, they should respect the influence of parental responsibility, at least to some extent (see, e.g., [Swift \(2005\)](#) for a defense of this position).

In many empirical applications, responsibility is unobserved, as it is here. In such cases, the equality of opportunity framework is usually operationalized using the identification axiom proposed by [Roemer \(1993\)](#), which states that the parents of two children who are at the same percentile of their type distribution of health have exercised identical responsibility.<sup>3</sup> Thus, if the cumulative distribution function of health for a type whose family participated in the program lies below the cumulative distribution function of health for the corresponding type who did not participate in the program, the type in the program needs less parental effort to obtain a particular level of child health than the type not in the program. If this holds for all levels of health, program participation unambiguously improves the opportunities for this type. Consequently, if the distribution of a type with treatment first-order stochastically dominates the distribution of the corresponding type that did not receive treatment, the program improves this type's opportunities. Similar reasoning applies to second-order stochastic dominance, with the caveat that second-order stochastic dominance can also be obtained by within-type, inequality-reducing transfers of health that do not fully respect the influence of parental responsibility.<sup>4</sup> Roemer's identification axiom does not necessarily imply that we would find children with and without treatment at exactly the same  $q$ th quantile (which is

---

<sup>2</sup>Race and educational background are circumstances because they should not influence the health opportunities parents can obtain for their children. Whether the family participates in the program largely determined by the by the locality in which they lived at the time the program began; therefore, this is outside of parental control.

<sup>3</sup>See [Roemer \(1993\)](#) and [Roemer \(1998\)](#) for a defense of this principle and [Fleurbaey \(1998\)](#)

<sup>4</sup>Fully respecting the influence of responsibility means that the health differences caused by parental responsibility are fully preserved by the program. Alternative notions of responsibility are weaker and require, for instance, that the program does not change the rank order of children's health. This weaker requirement is compatible with second order stochastic dominance.

the perfect positive quantile dependence found in Heckman, Jeffrey, and Clements (1997)); instead, it merely states that the comparison of the quantiles of the treated and corresponding untreated type is normatively relevant because it compares the health outcomes of children of parents who behaved equally responsibly.

Let  $F^C(h | c)$  denote the conditional distribution of children's health for parents with circumstances  $c$  in the control sample, and let  $F^T(h | c)$  denote the same distribution in the treatment sample. We say that the project improves the opportunities for the health of children with parental circumstances  $c$  if the conditional distribution  $F^T(h | c)$  first-order stochastically dominates the conditional distribution  $F^C(h | c)$ , and we test whether first-order stochastic dominance occurs. Thus, the issue of statistical inference arises. We follow Davidson and Duclos (2009), starting from nondominance as the null hypothesis. To illustrate the procedure for testing first order dominance and to describe the test more formally, let  $U \subseteq H$  be the union of the supports of  $F^C(h | c)$  and  $F^T(h | c)$ . We test the null hypothesis of nondominance of  $F^C(h | c)$  by  $F^T(h | c)$  :

$$\max_{z \in U} \left( F^T(z | c) - F^C(z | c) \right) \geq 0,$$

against the alternative hypothesis that  $F^T(h | c)$  first order stochastically dominates  $F^C(h | c)$  :

$$\max_{z \in U} \left( F^T(z | c) - F^C(z | c) \right) < 0.$$

This approach has the advantage of allowing us to draw the conclusion of dominance if we succeed in rejecting the null hypothesis; in other words, when the null is rejected, the only other possibility is dominance. By contrast, if dominance is the null hypothesis, as is the case in most empirical work to date, failure to reject dominance does not allow us to accept dominance. As Davidson and Duclos (2009) point out , taking nondominance as the null with continuous distributions comes at the cost that it is not possible to reject nondominance in favor of dominance over the entire support of the distribution.<sup>5</sup> Rejecting nondominance is normally possible only over restricted ranges of the observed variable. Thus, another merit of this approach is that it allows us to identify the maximal range over the supports of the distribution for which we are able to reject the null of nondominance and, therefore, to accept dominance in favor of the project. In this way, we can check whether we have dominance over ranges of the observed variable that are of special importance, such as the range below  $-2$  for

---

<sup>5</sup>Let  $\underline{h}$  be the lower bound of  $U$ . Evidently,  $F^T(\underline{h} | c) - F^C(\underline{h} | c) = 0$ , and so the maximum over  $U$  is never less than 0. Moreover, close to the boundaries of the support there may be too little information to reject non-dominance.

standardized height, which indicates stunting.

Of course, we must use the identical procedure to test the null of nondominance of  $F^T(h | c)$  by  $F^C(h | c)$  against the alternative hypothesis that  $F^C(h | c)$  dominates  $F^T(h | c)$ . If rejection occurs, we identify the maximal range over the support of the distribution for which we are able to reject the null of nondominance and to accept dominance against the project.<sup>6</sup> These issues are incorporated in the following weak version of improvements in opportunities which suffices for most of what we do in this paper.

**First Order Improvements:** The project leads to a first-order improvement of the opportunities of children with parental circumstances  $c$  if (i) there exists  $U^0 \subseteq U$  such that we can reject the null of nondominance of  $F^C(h | c)$  by  $F^T(h | c)$  against the alternative that  $F^T(h | c)$  dominates  $F^C(h | c)$  over  $U^0$ , and (ii) there exists no  $U^1 \subseteq U$  such that we can reject the null of nondominance of  $F^T(h | c)$  by  $F^C(h | c)$  against the alternative that  $F^C(h | c)$  dominates  $F^T(h | c)$  over  $U^1$ .

Assuming that the influence of parental responsibility on children's health need not be fully respected and that health is cardinally measurable, equalizing health outcomes within type becomes desirable such that, in case the project does not lead to a first order improvement, it becomes meaningful to ask whether the conditional distribution  $F^T(h | c)$  second order stochastically dominates the conditional distribution  $F^C(h | c)$ . Similar statistical issues as for first order stochastic dominance arise (see Davidson (2009)), leading to the following definition.

**Second Order Improvements:** The project leads to a second-order improvement of the opportunities of children with parental circumstances  $c$  if (i) the project does not lead to a first-order improvement, (ii) there exists  $U^0 \subseteq U$  such that we can reject the null of absence of second-order dominance of  $F^C(h | c)$  by  $F^T(h | c)$  against the alternative that  $F^T(h | c)$  second-order stochastically dominates  $F^C(h | c)$  over  $U^0$ , and (iii) there exists no  $U^1 \subseteq U$  such that we can reject the null of absence of second-order stochastic dominance of  $F^T(h | c)$  by  $F^C(h | c)$  against the alternative that  $F^C(h | c)$  second-order stochastically dominates  $F^T(h | c)$  over  $U^1$ .

Finally, when comparing conditional distribution functions to evaluate a program, it is important to note that inaccurate conclusions may be drawn when preprogram characteristics are not accounted for and when they differ for the treatment types in comparison with the control types (including compensation characteristics). Suppose we have two sets of characteristics, preprogram characteristics  $x \in X$ , which are not accounted for, and observable circumstances  $c$ . For the type with observed circumstances  $c_1$ , we then have

---

<sup>6</sup>Appendix 2.D contains more details about the stochastic dominance tests.

$$\begin{aligned}
F(h | c_1) &= \frac{\int_{\tilde{h}}^h f(\tilde{h}, c_1) d\tilde{h}}{f(c_1)} = \frac{\int_X \int_{\tilde{h}}^h f(\tilde{h}, c_1, x) d\tilde{h} dx}{f(c_1)} \\
&= \int_X \int_{\tilde{h}}^h f(\tilde{h} | c_1, x) \frac{f(c_1, x)}{f(c_1)} d\tilde{h} d\tilde{c}_2 = \int_X F(h | c_1, x) f(x | c_1) dx.
\end{aligned}$$

This equation clearly shows that the composition of the  $c_1$  type in terms of  $x$  matters. Indeed, suppose the treatment has no effect ( $F^C(h | c_1, x) = F^T(h | c_1, x)$ ), but the composition of those with circumstances  $c_1$  differs between the control and treatment types. Suppose that  $f^C(x | c_1)$  is higher than  $f^T(x | c_1)$  for favorable preprogram characteristics  $x$ , or characteristics for which  $F^C(h | c_1, x)$  is lower, and that  $f^C(x | c_1)$  is lower than  $f^T(x | c_1)$  for unfavorable preprogram characteristics. As a result,  $F^C(h | c_1)$  is smaller than  $F^T(h | c_1)$ , and we might erroneously infer that the treatment had an adverse effect on the opportunities of those with circumstances  $c_1$ .

## 2.3 Data description

In this section, we describe the Oportunidades program and the construction of treatment and control samples. We describe the selection of circumstances and outcomes and examine the data used to evaluate the program.

### 2.3.1 The Oportunidades program

The Oportunidades program is a conditional cash transfer program in which bimonthly cash transfers are provided to households in extreme poverty. The cash transfers are conditioned on the attendance of children in school, health care visits for all members of the household, and attendance at information sessions on primary health care and nutrition. Money for schooling constitutes the largest part of the conditional cash transfer. The total amount that a household receives depends on the number, age, and sex of its children. On average, households receive approximately 20 percent of their household consumption from such cash transfers.

Interventions for young children and their mothers are particularly emphasized. Prenatal and postpartum care visits, growth monitoring, immunization, and management of diarrhea and antiparasitic treatments are provided to mothers and young children. Children between the ages of 4 months and 23 months must have nine periodic medical check ups. From the age of 23 months until the child turns 19 years old, household members must have at least two check ups per year. Children between the ages of 6 and 23 months, lactating women and

low-weight children between the ages of 2 and 4 years receive milk-based and micronutrient fortified foods containing the daily recommended intake of zinc, iron, and essential vitamins.<sup>7</sup>

### 2.3.2 Sample Design

The selection of immediate and delayed treatment samples was undertaken in several steps (see, e.g., [INSP \(2005\)](#)). Highly deprived localities were identified by using a deprivation index computed on the basis of relevant sociodemographic data available from national censuses. Localities with at least 500 and not more than 2,500 inhabitants, that were categorized as having high or very high deprivation and that had access to an elementary school, a middle school and a health clinic were eligible for treatment. Localities were identified, and a random sample was constructed that was stratified by locality size. Within each state, localities were randomly assigned into treatment and control groups. A sample of 506 localities was finally selected for the study. A random procedure assigned 320 of these localities to receive immediate treatment; the remaining 186 began receiving treatment approximately 18 months later. In the selected localities, the poverty conditions of all households were evaluated, and households categorized as experiencing extreme poverty were included in the program. This categorization was based on household income, characteristics of the head of household, and variables related to dwelling conditions. Comments by a community assembly on the inclusion and exclusion of households were considered if they met certain criteria to identify beneficiary families. The randomized design enabled us to use the immediate treatment sample as the treatment group and the delayed treatment sample as the control group.<sup>8</sup> However, when we consider the effect of the program on the health outcomes of children between the ages of two and six years in 2003, most of these children grew up in families that were in the program for their entire lives. For children born before the delayed treatment began, this comparison can only show the effect of the difference in exposure when the children were young.<sup>9</sup> Therefore, and because we want to limit our study to an analysis of households that actually received cash transfers (this information is not available for the initial treatment sample), our treatment

---

<sup>7</sup>These supplements may also be given to children in households that are not receiving treatment (including children in the control sample) if signs of malnutrition are detected. This may lead to a downward bias of the estimated impact of Oportunidades (see also [Behrman, Parker, and Todd \(2009b\)](#), footnote 8).

<sup>8</sup>Most studies focus indeed on a comparison of the immediate and delayed treatment samples and thus evaluate the effect of differences in duration of program participation, see, e.g., [Schultz \(2004\)](#), [Behrman, Sengupta, and Todd \(2005\)](#) or [Behrman, Parker, and Todd \(2009a\)](#).

<sup>9</sup>In Appendix 2.G.3, we repeat the analysis for the children born after April 1998 (when the original treatment started) and before October 1999 (when delayed treatment started) taking the original treatment sample as treatment sample and the delayed treatment sample as control. The program effects are less outspoken, but some positive treatment effects remain - see also footnote 21.

sample is a subset of the delayed treatment sample.<sup>10</sup> Once the delayed treatment sample began receiving treatment, one had to construct a new control sample, with the intention of making it as similar as possible to the treatment samples (see, e.g., Todd (2004) and Behrman, Todd, Hernández, Urquieta, Attanasio, Angelucci, and Hernández (2006)). First, localities that did not meet the criteria for access to an elementary school, a middle school, and a health clinic were excluded. Next, a propensity score method was used that was based on data at the local level as a function of observed characteristics from the 2000 Census that permitted comparison with the localities of the original sample. This procedure led to a selection of 151 localities in which households that met the criteria for program eligibility were included in the control sample. We compare this control sample to the subset of the delayed treatment sample, as described above.

As we explained at the end of section 2.2, the households in the treatment and control samples must be comparable in terms of preprogram characteristics. There are important problems with the way the control sample was selected.<sup>11</sup> Matching at the local level was performed on the basis of a comparison with observable characteristics in 2000. By this time, the treatment sample had already received treatment. However, matching should have been performed on the basis of characteristics before treatment began. In addition, matching at the local level does not imply matching at the household level (see also Behrman and Todd (1999)). Moreover, we do not have data on all children of the households that were in the delayed treatment sample for three reasons (see table Appendix 2.A). First, some households dropped out of the sample because of sample attrition. Second, health data were only collected for a subsample of children. Third, because of problems with household identifiers, it was impossible to match all of the children for whom health data were available with only one household each. We only included unique matches in our samples (accounting for more than 80 percent of the children, fortunately). The second and third problems were also present in the control sample. As a result, the treatment and control samples may have differences in terms of preprogram characteristics.

For our empirical strategy in section 2.4, we first use a logistic regression approach to test whether there are statistically significant differences in composition between the treatment and control samples in 1997 for the households with children that were observed in

---

<sup>10</sup>Sensitivity analysis (reported in appendix 2.G.1) shows that the results are very similar when we compare the entire delayed treatment (including those for which no positive transfers were reported) and the control sample.

<sup>11</sup>This may explain why the control sample has rarely been used in academic papers. Recently, however, matched sampling was used to compare schooling (Behrman, Parker, and Todd (2009b) and Behrman and Parker (2010)) and work outcomes (Behrman and Parker (2010)) of the immediate treatment, delayed treatment and control samples.

2003.<sup>12</sup> We use a propensity score matching technique to match the four treatment types with the corresponding control types to correct for possible under- and overrepresentation of households with certain preprogram characteristics. This technique entails weighted sampling (see appendix 2.C). We compare the resulting weighted distributions at crucial points (such as standardized height smaller than  $-2$ , indicating stunting) to establish whether the treatment led to first- or second-order improvements of opportunities for each type by performing stochastic dominance tests on the weighted distribution functions.

### 2.3.3 Circumstances and outcomes

Ideally, normative theory requires us to obtain a full description of parental circumstances. In reality, an exhaustive description is not available from surveys, and the inclusion of an extensive set of circumstances is statistically unworkable for nonparametric procedures such as ours because of the limited number of observations. For these reasons, we limit ourselves to program participation and two additional circumstances.

The first circumstance refers to parental educational background. In the literature on equality of opportunity, this variable is used most frequently, is always statistically significant and has been shown to be the most important circumstance in Latin American countries (see, e.g., Bourguignon, Ferreira, and Menéndez (2007) and Ferreira and Gignoux (2011)). We measure educational background with a dichotomous variable indicating whether at least one parent completed primary education.<sup>13</sup> The second circumstance variable refers to parents' indigenous background. There is substantial literature indicating that indigenous people remain disadvantaged in Mexico (Olaiz, Rivera, Shamah, Rojas, Villalpando, and Hernández (2006); Psacharopoulos and Patrinos (1994); Rivera and Sepúlveda (2003); SEDESOL (2008)). We consider parents to have an indigenous background if at least one of them can speak or understand an indigenous language.

Combining these two binary characteristics with a binary characteristic indicating program participation yields eight types in Roemer's terminology. We partition the samples on the basis of parental indigenous origin (indigenous or nonindigenous) and parental level of education (primary or less than primary) to form the following types: indigenous, less than primary education (IL); indigenous, primary education (IP); nonindigenous, less than primary education (NL); nonindigenous, primary education (NP). Table 2.1 shows that there

<sup>12</sup>In 2003, in addition to the regular household data, an additional questionnaire with recall data was collected. The purpose of these retrospective questions was to compare the preprogram characteristics for the treatment samples with the new control sample.

<sup>13</sup>Appendix 2.G.2 reports the results when parental background is measured on the basis of mother's education only. The results are very similar to the ones we present here.

**Table 2.1:** Composition of the samples.

	Control sample		Treatment sample	
	#	%	#	%
All	1859	100	1125	100
IP	173	9.3	209	18.6
IL	241	13.0	274	24.4
NP	824	44.3	321	28.5
NL	621	33.4	321	28.5

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

are remarkable differences in the composition of the control sample and the treatment sample among these groups. Clearly, the control sample contains fewer indigenous children and more nonindigenous children with at least one parent who completed primary education than the treatment sample. Because we are comparing cumulative distribution functions of types in the control sample with the corresponding types in the treatment sample, this creates no problem for our analysis. However, as shown in section 2.2, problems arise when there are important differences in terms of preprogram characteristics between the treatment and control types that are compared.

We focus on several health outcomes. Two important measures of malnutrition for children are anemia, which is defined as hemoglobin levels lower than 11 grams per deciliter, and stunting, which covers a wider range of nutritional deficiencies and is defined as height for age below  $-2$  standard deviations of the WHO International Growth Reference. The latter implies that in a reference population, approximately 2.3 percent of the population is stunted. As reviewed by [Grantham-McGregor and Ani \(2001\)](#), anemia (iron deficiency) in infancy has been associated with poorer cognition, school achievement, and behavioral problems into middle childhood. [Branca and Ferrari \(2002\)](#) point out that stunting is associated with developmental delay, delayed achievement of developmental milestones (such as walking), later deficiencies in cognitive ability, reduced school performance, increased child morbidity and mortality, higher risk of developing chronic diseases, impaired fat oxidation (stimulating the development of obesity), small stature later in life, and reduced productivity and chronic poverty in adulthood. In addition to actual stunting, height has a positive effect on completed years of schooling, earnings (see, e.g., [Alderman, Hoddinott, and Kinsey \(2006\)](#)), and cognitive and noncognitive abilities (see, e.g., [Case and Paxson \(2008\)](#) and [Schick and Steckel \(2010\)](#)) throughout the distribution. Therefore, we treat our two measures of malnutrition as dichotomous and continuous variables, focusing on the fraction of anemic (stunted)



**Table 2.2:** Health outcomes of 2-6 year old children in 2003.

(a) Control sample							
	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.24	12.0	0.32	-1.46	0.24	0.58	0.17
IP	0.36	11.6	0.50	-1.99	0.23	0.57	0.19
IL	0.30	11.9	0.64	-2.40	0.30	0.64	0.13
NP	0.18	12.2	0.20	-1.13	0.22	0.56	0.18
NL	0.25	12.0	0.32	-1.47	0.25	0.58	0.18

(b) Treatment sample							
	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.23	12.1	0.34	-1.58	0.20	0.67	0.12
IP	0.27	12.0	0.35	-1.63	0.14	0.64	0.14
IL	0.29	11.7	0.43	-1.82	0.16	0.72	0.11
NP	0.13	12.5	0.26	-1.32	0.24	0.68	0.10
NL	0.24	12.2	0.33	-1.58	0.22	0.63	0.16

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

children and on the entire distribution of hemoglobin levels (standardized height). Another health outcome is based on the standardized Body Mass Index (BMI); children are at risk of being overweight if their standardized BMI is larger than 1.15.<sup>14</sup> In a reference population, this cutoff value indicates that 15 percent of children are at risk of being overweight. Overweight children have delayed skill acquisition at young ages (Cawley and Spiess, 2008), are more likely to have psychological or psychiatric problems, have increased cardiovascular risk factors, have increased incidence of asthma and diabetes ((Reilly, Metheven, McDowell, Hacking, Alexander, Stewart, and Kelnar, 2003)), are more likely to be obese as adults ((Serdula, Ivery, Coates, Freedman, Williamson, and Byers, 1993)), and may earn lower wages ((Cawley, 2004)). A final health outcome is based on the number of days parents reported that the child was sick during the previous four-week period. We consider the percentage of children reporting zero days and more than three days. Table 2.2 provides information on the outcome variables of the control and treatment samples.

Considering all households, it is striking that the different entries are similar for all health outcomes in the control and treatment samples, with the exception of the number of days sick; fewer sick days were reported for children in the treatment sample than in the control sample. Approximately one child in four is anemic, and one in three is stunted. Compared

<sup>14</sup>The incidence of underweightness is lower than in a reference population.

with the reference population, our sample contains far too many stunted children and too many children at risk of being overweight.

Interesting but predictable patterns emerge when considering the distribution of health outcomes over the types.<sup>15</sup> Comparing the IL type with the NL type and the IP type with the NP type, indigenous children have worse health outcomes than nonindigenous children, except for the risk of being overweight in the treatment sample. The differences are substantial, particularly for hemoglobin concentration and standardized height in the control sample. Comparing the IL type with the IP type and the NL type with the NP type, the differences between children who had at least one parent who completed primary education and children whose parents had less than primary education are less obvious. The largest differences occur for standardized height; here having a parent who completed primary education is a clear advantage. Overall, these results are in line with the previous literature (see, e.g., Backstrand, Allen, Pelto, and Chávez (1997); Fernald and Neufeld (2006); González de Cossío, Rivera, González-Castell, and Monterrubio (2009); Rivera and Sepúlveda (2003); Rivera, Monterrubio, González-Cossío, García-Feregrino, García-Guerra, and Sepúlveda (2003)).

## 2.4 Empirical Results

We now use the data described in the previous section to evaluate the Oportunidades program. We show that the treatment and control samples are not comparable in terms of preprogram characteristics, and we apply a propensity score matching technique to make them comparable. We apply the methodology presented in section 2.2 on the resulting samples to evaluate the program. We then compare the results to previous studies.

### 2.4.1 Comparison of weighted treatment and control types

As stated at the end of section 2.2, a crucial assumption in the identification of treatment effects on the basis of a simple comparison of the outcomes of treatment and control samples is that  $f^C(x | c_1) = f^T(x | c_1)$ , implying that the two samples must be similar in terms of preprogram characteristics. If that is the case, after conditioning on  $c_1$ , observing  $x$  does not provide any information about whether an observation belongs to the treatment or control sample. We test this hypothesis as described below.

We construct a sample containing members of both the control and treatment samples. Next, we perform a logistic regression in which the dependent variable takes the value one if

---

<sup>15</sup>The types might differ in terms of characteristics that do not enter the definition of type and in terms of preprogram characteristics.

the observation belongs to the control sample and the value zero if it belongs to the treatment sample.

Explanatory variables are characteristics of the family, characteristics of the family's dwelling, family assets, and state of residence (see appendix 2.B for more details). These characteristics were measured in 1997, before the program started.<sup>16</sup> The results are reported in table 2.B.1 in appendix 2.B. We find that many of the characteristics significantly affect the probability that the observation comes from the control sample, indicating that the hypothesis that treatment and control samples are comparable in terms of the composition of their preprogram characteristics must be rejected.

In the identification of average treatment effects, a standard way to address differences in the composition of the treatment and control samples is to use propensity score matching techniques. The goal is to make the treatment and control samples more comparable by weighting different observations based on the estimated probability that the observation belongs to the control sample, as determined by the logistic regression discussed in the previous paragraph. Appendix 2.C explains this procedure and how the weighting is used to obtain estimates of the relevant distribution functions. The weighting procedure has a substantial effect on the Roemer motivation for considering cumulative distribution functions (Roemer's identification axiom), as we discuss in appendix 2.E.<sup>17</sup> Appendix 2.F provides the equivalent of table 2.2 for the weighted (matched) samples.

In table 2.3, we use the weighted samples to consider the effect of the treatments on the fraction of children who are anemic, stunted, or at risk of being overweight. We use the same samples to examine the fraction of children for whom zero sick days or more than three sick days during the previous four weeks were reported. Effects that are statistically significantly different from zero at the 5 percent level of significance are indicated by "\*\*," and effects that are statistically significantly different from zero at the 10 percent level of significance are indicated by one "\*". Each entry provides the effect of the treatment. From an opportunity perspective, a desirable effect on these fractions indicates that less responsibility allows parents to prevent their children from being anemic, stunted, at risk of being overweight, sick, or more than three days sick in the previous four-week period.

We see that the treatment effects reported in table 2.3 are substantial, and all significant effects of the program are in a desirable direction. For each health indicator, we find at

<sup>16</sup>For the control sample this is based on recall data -see also footnote 12.

<sup>17</sup>Because health is also influenced by preprogram characteristics, we can no longer infer from the percentile in the distribution of health for each type the corresponding responsibility; the same percentile will be obtained by people with different combinations of responsibility and preprogram characteristics. In the appendix we show that, under certain assumptions, the weighting procedure guarantees that those that are at the same percentile in the weighted treatment and the control sample have the same expected responsibility.

**Table 2.3:** Difference between control and treatment in fraction of anemic, stunted, at risk of being overweight and days sick. Weighted samples.

	Anemic	Stunted	Risk Overweight	0 Days Sick	> 3 Days Sick
All	-0.03	0.01	-0.04	0.09**	-0.06**
IP	-0.17**	-0.17**	-0.08	0.09	-0.06
IL	-0.05	-0.18*	-0.11**	0.10*	-0.05*
NP	-0.08**	0.05	0.03	0.07	-0.09**
NL	0.00	-0.01	-0.04	0.06	-0.02

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education. One (two) "\*" indicates that the effect is statistically significant from zero at the ten (five) percent level. Standard errors corrected for clustering at locality level.

least one significant desirable treatment effect for one of the types. The table suggests that the program works well, particularly for children of indigenous origin without a parent who completed primary education. This type is likely to be the most disadvantaged, as table 2.2 suggests.

Children of indigenous origin with a parent who completed primary education have an improvement in all indicators, although the effects are only significant for the fraction of anemic and stunted children. For nonindigenous children, the results are less obvious. The fraction of nonindigenous children who are anemic decreases because of the program, but the results presented in table 2.3 identify no other significant treatment effects for nonindigenous children.

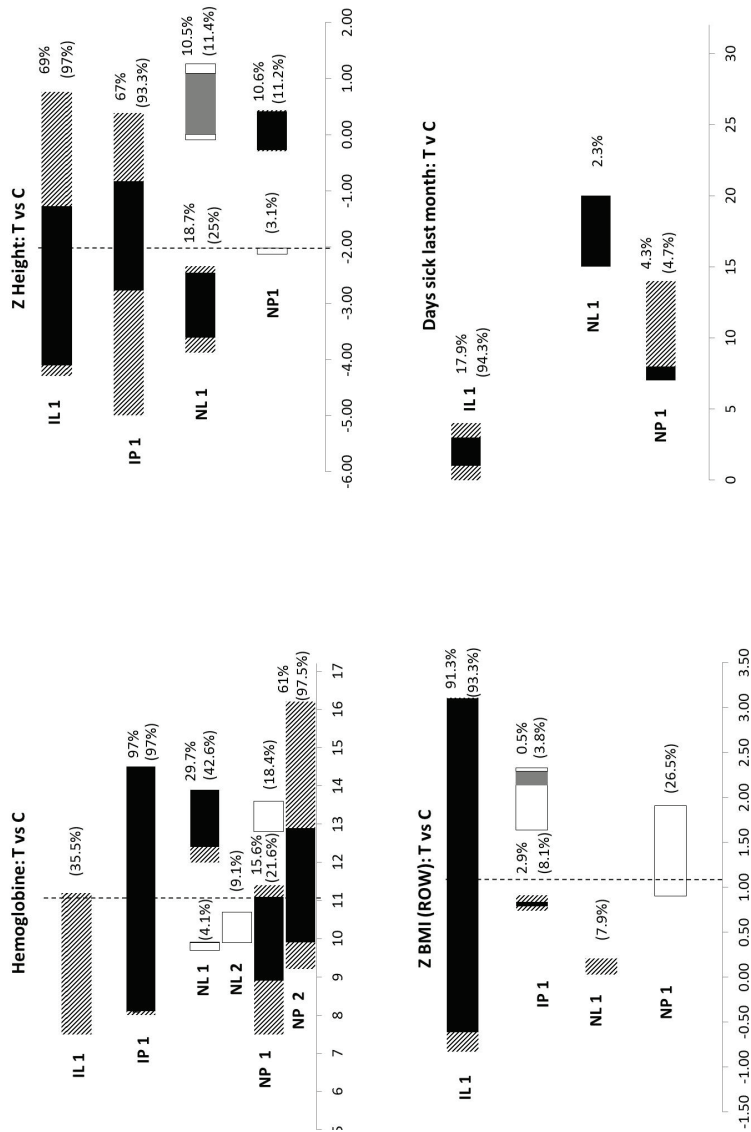
Figure 2.1 presents the results of the stochastic dominance tests, using the procedure explained in section 2.3.<sup>18</sup> The horizontal axis denotes the numerical value of the variable of interest (hemoglobin concentration, standardized height, standardized BMI, and reported days sick).

The black (grey) boxes depict the maximal range over the support of the distributions for which the null of nondominance is rejected at the 5 percent level of significance in favor of a desirable (undesirable) effect of the treatment. Hatched (white) boxes indicate the same at a significance level of 10 percent. When hatched (white) boxes are adjacent to a black (grey) box, they show how far the rejection range of the null can be extended for the 10 percent level of significance. Each row contains an acronym "XYi" of which the first two characters, "XY", indicate the name of the types that are compared (XY = IL, IP, NL, or

<sup>18</sup>Because of many zero observations, this test procedure cannot be used for the number of days sick. Here, the stochastic dominance test is based on a standard test for the difference between the cumulative distribution functions at the natural numbers between 0 and 30. The intervals shown for this health outcome connect the points in the support where the difference between the cumulative distribution functions is statistically significant.

NP), and the character “i” indicates whether the test refers to first- ( $i = 1$ ) or second- ( $i = 2$ ) order stochastic dominance. The numbers in parentheses behind the boxes show the percentage of observations of the treated type within the black or grey (hatched or white) box.

Figure 2.1: Stochastic dominance results.



For example, in the top left panel of Figure 2.1, the hatched box labeled “IL” shows that, using a 10 percent level of significance, the null hypothesis that the cumulative distribution of the treatment type does not first-order stochastically dominate the distribution of the control type must be rejected against the alternative, that the distribution of the treatment type first-order stochastically dominates the distribution of the control type over the range [7.5, 11.2], which contains 35.5 percent of the treated type. The hypothesis of nondominance can only be rejected at the 10 percent level of significance. Thus, we tested the null hypothesis of the absence of second-order stochastic dominance in favor of the treatment against the alternative, that the distribution of the treatment type second-order stochastically dominates the distribution of the control type at the 5 percent level of significance. We failed to reject the null, such that no box “IL” is drawn. For IP types, the black box labeled “IP1” indicates that the null hypothesis of nondominance can be rejected at the 5 percent level of significance over the range [8.1, 14.5], which contains 97 percent of the treated IP type. When we increase the level of significance to 10 percent, the hatched box shows that the rejection interval enlarges only marginally, to [8.0, 14.5]. For NL types, when testing for first-order stochastic dominance, we find a white box over the small range of [9.7, 9.9] with very few observations of the treatment type and a solid black box further up in the distribution. When testing NL types for second-order stochastic dominance, we find a small white box. On balance, the evidence for this type against treatment is not strong. Finally, for NP types, we have first a solid black and then a white box. The latter is only significant at the 10 percent level of significance and occurs at a less important part of the distribution (above 11, when children are no longer anemic). When testing for second-order stochastic dominance, we see a solid black box labeled “NP2,” indicating that the project leads to second-order improvement,<sup>19</sup> and this type is also positively affected by the program.

The other panels in Figure 2.1 can be similarly interpreted. In the top right panel, we see that the treatment leads to first-order improvements in the standardized height for IL and IP types over large and crucial parts of the support (standardized height below -2). For NL types, we find a first-order stochastic dominance effect in favor of the treatment in an important part of the distribution (standardized height below -2) and an adverse effect higher

---

<sup>19</sup>Observe that the “NP1” interval is not a subset of the “NP2” interval. This is because the test procedure for first (second) order stochastic dominance identifies the point in the support where the difference between the (cumulated) cumulative distribution functions is most significant, and then constructs the interval around this point. There is no reason why the point (and hence the intervals) identified should be the same, or why the intervals should be related by set inclusion. Moreover, first order stochastic dominance over a particular interval does not imply second order stochastic dominance over that same interval since for second order dominance the values of the cumulative distribution functions to the left of the first interval are also relevant. Hence it can even occur that we find an interval over which we reject non first order stochastic dominance but cannot find an interval over which we reject non second order stochastic dominance.

up in the distribution. There is evidence of a marginal perverse first-order treatment effect at a significance level of 10 percent on standardized height for NP types over a small range of  $[-2.11, -2.00]$ , which contains only 3 percent of the observations of the treated type, and a positive effect higher up in the distribution. No second-order stochastic dominance effects can be established for the nonindigenous types. In the bottom left panel, we concentrate on what occurs at the right of the dotted vertical line, which represents children at risk of being overweight. We see positive, first-order stochastic dominance effects at the 5 percent level of significance for IL types and some evidence of marginally significant perverse treatment effects for IP and NP types. The bottom right panel shows first-order improvements for IL, NL, and NP types. The intervals reported here, except for IL, contain few observations, because of the high frequency of zero reported sick days (see table 2.2).

The results reported in table 2.3 and figure 2.1 are consistent. The stochastic dominance results provide more detail and identify effects in important parts of the distribution that would otherwise go unnoticed, such as the positive first-order stochastic dominance effect on standardized height for NL children. If first-order improvements cannot be found and the influence of parental responsibility is not to be fully respected, then second-order stochastic dominance provides a way to determine whether the program has positive effects. Second-order improvements occur only once in our application, for the hemoglobin concentration of NP types. In summary, we find strong evidence of positive treatment effects for children of indigenous origin, particularly for those without a parent who completed primary education. The evidence for children from nonindigenous origin is not as strong, but enrollment in the program also seems to have positive effects on health opportunities for these children, on balance.

#### 2.4.2 Comparison to previous studies

Diaz and Handa (2006) use propensity score matching techniques to construct alternative control samples from the Mexican national household survey. They compute average treatment effects by comparing the immediate treatment sample after eight months of receiving program benefits with the delayed treatment sample (who had not yet received benefits), on the one hand, and their newly constructed control samples, on the other. They conclude, "The PSM [propensity score matching] technique requires an extremely rich set of covariates, detailed knowledge of the beneficiary selection process, and the outcomes of interest need to be measured as comparably as possible in order to produce viable estimates of impact" (p.341). In our case, the outcomes are measured in identical ways in the delayed treatment and control samples, and the control sample is constructed following the beneficiary selection process as



closely as possible. Our selection of covariates for the propensity score matching closely follows [Behrman, Parker, and Todd \(2009b\)](#), who use almost identical covariates in comparing the effects on schooling outcomes of the short-run differential exposure (between the immediate and delayed treatment samples) with the long-run differential exposure (between the immediate treatment and control samples). They find that longer exposure produces larger effects, and the differences between the order of magnitude of the short- and long-run effects are reasonable. This finding suggests that the propensity score matching technique we use can produce reliable estimates of average treatment effects.

The interpretation of the difference between the distributions of the weighted treatment and control samples as a treatment effect depends on the extent to which the weighting procedure manages to correct for possibly unobserved heterogeneity caused by the imperfect randomness of the assignment to treatment and control groups. Of course, it is not possible to test this directly, but we can compare our results to the findings in the literature that consider differences in children’s health outcomes between immediate and delayed treatment samples. [Rivera, Sotres-Alvarez, Habicht, Shamah, and Villalpando \(2004\)](#) compare the health outcomes of children younger than 12 months old in 1997. They find that in 1999 after 12 months of treatment, children in the immediate treatment sample had higher mean hemoglobin values than the children from the delayed treatment sample, who were untreated up to that point. After the immediate treatment sample had received 24 months of treatment and the delayed treatment sample had received approximately six months of treatment, children from the immediate treatment sample had grown more than children in the delayed treatment sample, and the differences in height were significantly larger for households with low socioeconomic status (a score based on dwelling characteristics, possession of durable goods, and access to water and sanitation). [Gertler \(2004\)](#) finds similar results for children aged 0 to 35 months in 1997, stating that “treatment children were 25.3 percent less likely to be anemic and grew about 1 centimeter more during the first year of the program” (p. 340). Both of these differences are statistically significant at the 1 percent level. Unfortunately, Gertler does not report whether the effect differs for different subgroups, such as our types. Hemoglobin levels, unlike height, were not observed before the program started. Therefore, the results for hemoglobin levels do not control for child fixed effects as opposed to growth effects, as noted by [Behrman and Hoddinott \(2005\)](#). They investigate the effect on the height of children who were between 4 and 48 months of age when treatment began in August 1998. They find that when child fixed effects are not included, treatment has a significant negative effect on child height for children between 4 and 36 months of age. However, if child fixed effects are controlled (by considering the difference between 1999 and 1998), the treatment

effect becomes significantly positive at approximately one centimeter, as in [Gertler \(2004\)](#).<sup>20</sup> Notably, program effects are larger for children in households in which the head of the household speaks an indigenous language and the mother is more educated.<sup>21</sup>

Finally, [Fernald, Gertler, and Neufeld \(2008\)](#) use a different approach. They combine the data of both the immediate and delayed treatment samples to estimate the effect of the size of the conditional cash transfer received on children between 24 and 68 months of age in 2003, when the children's height was measured. Increasing the size of the transfer leads to higher height-for-age scores, a lower prevalence of stunting and a lower prevalence of obesity. Parental level of education and whether the head of the household spoke an indigenous language were not significant controls in their model.

Overall, these findings are in line with ours. The program has significant positive effects on children's height and hemoglobin concentration levels. Larger effects tend to be found for households in which an indigenous language is spoken. This finding is compatible with [Fernald, Gertler, and Neufeld \(2008\)](#) because, in general, indigenous families receive larger cash transfers than nonindigenous families based on the finding that they tend to have more children. Our results visualize where in the distribution the program is most effective for different types, and we can see that the program is most powerful for the most disadvantaged types, children of indigenous origin.

## 2.5 Conclusion

There is a growing body of literature on the measurement of inequality of opportunity (for an overview, see, e.g., [Ramos and Van de gaer \(2015\)](#)). Thus far, the ideas in the literature have not been applied to evaluate social programs. We propose a methodology to do so.

We bring together insights from the literature on equality of opportunity, the literature on program evaluation, and the literature on testing for stochastic dominance. [Roemer \(1993\)](#)'s normative approach to equality of opportunity indicates that we should focus on types and that, if responsibility characteristics are unobserved, individuals at the same percentile of the

---

<sup>20</sup>[Behrman and Hoddinott \(2005\)](#) obtain the same pattern when considering standardized height-for-age scores.

<sup>21</sup>We compare the health outcomes of immediate and delayed treatment in the appendix 2.G.3 of the paper for children born between the beginning of the initial and the beginning of the delayed treatment. This substantially limits the size of the sample. Moreover, because all of these children received at least three years of treatment by the time their health outcomes were measured, few significant effects can be found, particularly for hemoglobin concentration and reported days sick. This indicates that these variables are more sensitive to nutritional status in the immediate past than in the more distant past. We find significant positive effect on standardized height for indigenous children without parental primary education over a large range of the support of the distribution and for nonindigenous children with parental primary education over a limited support of the distribution. Again, the evidence is in favor of the program.

distribution of the outcome within their type have exercised a comparable degree of responsibility. This approach provides a normative foundation for the comparison of cumulative distribution functions of corresponding treatment and control types. The literature on program evaluation stresses that care should be taken to ensure that the treatment and control samples are comparable in terms of preprogram characteristics. If they are not, propensity score matching techniques can be used to make the samples more comparable. Hence, we test whether the treatment and control samples are comparable in terms of preprogram characteristics and since the test fails, we propose a weighted sampling method based on standard propensity score matching techniques to make treatment and control types comparable. Finally, Davidson and Duclos (2009) and Davidson (2009) propose a new technique to test for stochastic dominance, taking nondominance as the null so that rejection of the null implies dominance. Their test procedure is particularly suited to our study because it allows us to see where dominance can be established along the distribution.

We applied our procedure to study the effect of the Mexican Oportunidades program on children's health opportunities. We can draw two conclusions about the proposed methodology. First, in our application (as in the applications by Lefranc, Pistolesi, and Trannoy (2008), Lefranc, Pistolesi, and Trannoy (2009), Peragine and Serlenga (2008), and Rosa Dias (2009)), looking for second-order stochastic dominance does not significantly add to the conclusions drawn from first-order stochastic dominance. Thus, whether the influence of parental responsibility is to be fully respected does not substantially affect the conclusions. Second, the treatment and control samples differed substantially in terms of preprogram characteristics. Therefore, it is important to use weighted sampling based on techniques such as propensity score matching to make the samples (more) comparable. Concerning the actual effects of the program, our results indicate that the Oportunidades program has a substantially favorable effect on the health opportunities of the most disadvantaged children, that is, those with parents of indigenous origin and without a parent who completed primary education. Additionally, the effects on children of indigenous origin with a parent who completed primary education are sizable and important. The effects on nonindigenous children are less obvious, but the overall evidence in this paper indicates that the program also results in better health opportunities for these children.



## Bibliography

- ALDERMAN, H., J. HODDINOTT, AND B. KINSEY (2006): “Long term consequences of early childhood malnutrition,” *Oxford Economic Papers*, 58(3), 450–474.
- ASLAM, M., AND G. KINGDON (2010): “Parental Education and Child Health-Understanding the pathways of impact in Pakistan,” *RECOUP Working Paper No. 30*.
- BACKSTRAND, J., L. ALLEN, G. PELTO, AND A. CHÁVEZ (1997): “Examining the gender gap in nutrition: An example from rural Mexico,” *Social Science & Medicine*, 44(11), 1751–1759.
- BEHRMAN, J., AND J. HODDINOTT (2005): “Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition,” *Oxford Bulletin of Economics and Statistics*, 67(4), 547–569.
- BEHRMAN, J., AND S. PARKER (2010): “Do conditional cash transfers generate lasting benefits? a five year follow-up of PROGRESA/Oportunidades,” *mimeo*.
- BEHRMAN, J., S. PARKER, AND P. TODD (2009a): “Medium-term impact of the oportunidades conditional cash transfer program on rural youth in Mexico,” in *Poverty, Inequality and Poverty in Latin America*, ed. by S. Klasen, and N. F, pp. 219–270. MIT Press, Cambridge.
- (2009b): “Schooling Impacts of Conditional Cash Transfers on Young Children: Evidence from Mexico,” *Economic Development and Cultural Change*, 57(3), 439–477.
- BEHRMAN, J., P. SENGUPTA, AND P. TODD (2005): “Progressing through PROGRESA: An Impact Assessment of a School Subsidy Experiment in Rural Mexico,” *Economic Development and Cultural Change*, 54(1), 237–275.
- BEHRMAN, J., AND P. TODD (1999): “Randomness in the experimental samples of PROGRESA –Education, Health, and Nutrition Program,” *International Food Policy Research Institute*.
- BEHRMAN, J., P. TODD, B. HERNÁNDEZ, J. URQUIETA, O. ATTANASIO, M. ANGELUCCI, AND M. HERNÁNDEZ (2006): *Evaluación externa de impacto del programa Oportunidades 2006*. Instituto Nacional de Salud Pública.

- BLACK, S., P. DEVEREUX, AND K. SALVANES (2007): "From the Cradle to the Labor Market? The Effect of Birth Weight on Adult Outcomes," *The Quarterly Journal of Economics*, 122(1), 409–439.
- BOSSERT, W. (1995): "Redistribution mechanisms based on individual characteristics," *Mathematical Social Sciences*, 29(1), 1–17.
- BOURGUIGNON, F., F. FERREIRA, AND M. MENÉNDEZ (2007): "Inequality of opportunity in Brazil," *Review of Income and Wealth*, 53(4), 585–618.
- BRANCA, F., AND M. FERRARI (2002): "Impact of Micronutrient Deficiencies on Growth: The Stunting Syndrome," *Annals of Nutrition and Metabolism*, 46(Suppl. 1), 8–17.
- BREIEROVA, L., AND E. DUFLO (2003): "The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers?," *OECD Development Centre Working Papers. OECD Publishing*, 217.
- CASE, A., AND C. PAXSON (2008): "Stature and Status: Height, Ability and Labor Market Outcomes," *Journal of Political Economy*, 116(3), 499–532.
- CAWLEY, J. (2004): "The impact of obesity on wages," *Journal of Human Resources*, 39(2), 451–474.
- CAWLEY, J., AND C. SPIESS (2008): "Obesity and skill attainment in early childhood," *Economics and Human Biology*, 6, 388–397.
- CHEN, W., AND J. DUCLOS (2008): "Testing for Poverty Dominance: an Application to Canada," *IZA Discussion Paper N 2829*.
- DAVIDSON, R. (2009): "Testing for Restricted Stochastic Dominance: Some Further Results," *Review of Economic Analysis*, 1(1), 34–59.
- DAVIDSON, R., AND J. DUCLOS (2009): "Testing for Restricted Stochastic Dominance," *GREQAM Document de Travail 2009-38*, (06-09).
- DE BARROS, R., F. FERREIRA, J. MOLINAS VEGA, AND J. CHANDUVI (2009): *Measuring Inequality of Opportunities in Latin America and the Caribbean*. The World Bank.
- DESAI, D., AND S. ALVA (1998): "Maternal education and child health: Is there a strong causal relationship?," *Demography*, 35(1), 71–81.

- DIAZ, J., AND S. HANDA (2006): "An Assessment of Propensity Score Matching as a Non-experimental Impact Estimator: Evidence from Mexico's PROGRESA Program," *Journal of Human Resources*, 41(2), 319–345.
- FERNALD, L., P. GERTLER, AND L. NEUFELD (2008): "Role of cash in Conditional Cash Transfer programmes for child health, growth, and development: an analysis of Mexico's Oportunidades," *The Lancet*, 371(9615), 828–837.
- FERNALD, L., AND L. NEUFELD (2006): "Overweight with concurrent stunting in very young children from rural Mexico: prevalence and associated factors," *European Journal of Clinical Nutrition*, 61(5), 623–632.
- FERREIRA, F. H. G., AND J. GIGNOUX (2011): "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America," *Review of Income and Wealth*, forthcoming.
- FISZBEIN, A., N. R. SCHADY, AND F. H. FERREIRA (2009): *Conditional cash transfers: reducing present and future poverty*. World Bank Publications.
- FLEURBAEY, M. (1995): "The requisites of equal opportunity," in *Social Choice, Welfare and Ethics*, ed. by M. Salles, and N. Schofield, pp. 37–53. Cambridge University Press.
- (1998): "Equality among responsible individuals," in *Freedom in Economics: New Perspectives in normative economics*, ed. by J. Laslier, M. Fleurbaey, N. Gravel, and A. Trannoy, pp. 206–234. Routledge, London.
- (2008): *Fairness, Responsibility and Welfare*. Oxford University Press, Oxford.
- FOSTER, J., J. GREER, AND E. THORBEKE (1984): "A Class of Decomposable Poverty Measures," *Econometrica*, 52(3), 761–766.
- GERTLER, P. (2004): "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment," *American Economic Review*, 94(2), 336–341.
- GONZÁLEZ DE COSSÍO, T., J. RIVERA, D. GONZÁLEZ-CASTELL, AND E. MONTEERRUBIO (2009): "Child malnutrition in Mexico in the last two decades: prevalence using the new WHO 2006 growth standards," *Salud Pública de México*, 51.
- GRANTHAM-MCGREGOR, S., AND C. ANI (2001): "A Review of Studies on the Effect of Iron Deficiency on Cognitive Development in Children," *The Journal of Nutrition*, 131(2), 649S–668S.

- HECKMAN, J. (1992): "Randomization and Social Policy Evaluation," in *Evaluating Welfare and Training Programs*, ed. by C. Manski, and I. Garfinkel, pp. 201–230. Cambridge: Harvard University Press.
- HECKMAN, J., S. JEFFREY, AND N. CLEMENTS (1997): "Making the Most Out Of Programme Evaluations and Social Experiments: Accounting For Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64(4), 487–535.
- INSP (2005): "General Rural Methodology Note," *Instituto Nacional de Salud Pública*, INSP2005.
- LEFRANC, A., N. PISTOLESI, AND A. TRANNOY (2008): "Inequality of Opportunities Vs. Inequality of Outcomes: Are Western Societies All Alike?," *Review of Income and Wealth*, 54(4), 513–546.
- (2009): "Equality of opportunity and luck: Definitions and testable conditions, with an application to income in France," *Journal of Public Economics*, 93(11-12), 1189–1207.
- NASCHOLD, F., AND C. BARRETT (2010): "A Stochastic Dominance Approach to Program Evaluation with an Application to Child Nutritional Status in Kenya," Working Paper.
- OLAIZ, G., J. RIVERA, T. SHAMAH, R. ROJAS, S. VILLALPANDO, AND M. HERNÁNDEZ (2006): "Encuesta Nacional de Salud y Nutrición 2006 [National Health and Nutrition Survey 2006]," *Instituto Nacional de Salud Pública*.
- O'NEILL, D., O. SWEETMAN, AND D. VAN DE GAER (2000): "Equality of Opportunity and Kernel Density Estimation: An application to Intergenerational Mobility," in *Advances in Econometrics*, ed. by T. Fomby, and R. C. Hill, vol. 14, pp. 259–274. JAI Press, Stanford.
- PARKER, S., L. RUBALCAVA, AND G. TERUEL (2008): "Evaluating Conditional Schooling and Health Programs," in *Handbook of Development Economics*, ed. by T. Schultz, and J. Strauss, vol. 4, pp. 3963–4035. Elsevier.
- PERAGINE, V., AND L. SERLENGA (2008): "Higher education and equality of opportunity in Italy," in *Inequality of opportunity: papers from the Second ECINEQ Society Meeting, Research on Economic Inequality*, ed. by J. Bishop, and B. Zheng, vol. 16, pp. 67–97. Emerald Group Publishing, Bingley.
- PSACHAROPOULOS, G., AND H. PATRINOS (1994): "Indigenous People and Poverty in Latin America," *The World Bank*.



- RAMOS, X., AND D. VAN DE GAER (2015): "Approaches to inequality of opportunity: Principles, measures, and evidence," *Journal of Economic Surveys*.
- REILLY, J., E. METHEVEN, Z. MCDOWELL, B. HACKING, D. ALEXANDER, L. STEWART, AND C. KELNAR (2003): "Health consequences of obesity," *Archives of disease in childhood*, 88(9), 748–752.
- RIVERA, J., E. MONTEERRUBIO, T. GONZÁLEZ-COSSÍO, A. GARCÍA-FEREGRINO, A. GARCÍA-GUERRA, AND J. SEPÚLVEDA (2003): "Nutritional status of indigenous children younger than five years of age in Mexico: Results of a national probabilistic survey.," *Salud Pública de México*, 45, S466–S476.
- RIVERA, J., AND J. SEPÚLVEDA (2003): "Conclusions from the Mexican National Nutrition Survey 1999: Translating results into nutrition policy," *Salud Pública de México*, 45, S565–S575.
- RIVERA, J., D. SOTRES-ALVAREZ, J. HABICHT, T. SHAMAH, AND S. VILLALPANDO (2004): "Impact of the Mexican Program for Education, Health, and Nutrition (PROGRESA) on Rates of Growth and Anemia in Infants and Young Children," *The Journal of the American Medical Association*, 291(21), 2563–2570.
- ROEMER, J. (1993): "A Pragmatic Theory of Responsibility for the Egalitarian Planner," *Philosophy & Public Affairs*, 22(2), 146–166.
- (1998): *Equality of opportunity*. Harvard University Press, Cambridge MA.
- ROSA DIAS, P. (2009): "Inequality of opportunity in health: evidence from a UK cohort study," *Health Economics*, 18(9), 1057–1074.
- SCHICK, A., AND R. STECKEL (2010): "Height as a Proxy for Cognitive and Non-Cognitive Ability," *NBER Working Paper N 16570*.
- SCHULTZ, P. (2004): "School subsidies for the poor: evaluating the Mexican Progresa poverty program," *Journal of Development Economics*, 74(1), 199–250.
- SEDESOL (2008): *Evaluación externa del Programa Oportunidades 2008. A diez años de intervención en zonas rurales (1997-2007)*. Ministry of Social Development of Mexico (SEDESOL).
- SERDULA, M., D. IVERY, R. COATES, D. FREEDMAN, D. WILLIAMSON, AND T. BYERS (1993): "Do obese children become obese adults? A Review of the literature," *Preventive Medicine*, 22, 167–177.

- SILVERMAN, B. W. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.
- SWIFT, A. (2005): "Justice, luck, and the family: the intergenerational transmission of economic advantage from a normative perspective," in *Unequal chances: family background and economic success*, ed. by S. Bowles, H. Gintis, and M. Osborne Groves, pp. 256–276. Princeton University Press.
- TODD, P. (2004): "Design of the evaluation and method used to select comparison group localities for the six year follow-up evaluation of Oportunidades in rural areas," Discussion paper, International Food Policy Research Institute.
- TRANNOY, A., T. SANDY, F. JUSOT, AND M. DEVAUX (2010): "Inequality of opportunities in health in France: a first pass," *Health Economics*, 19(8), 921–938.
- VERME, P. (2010): "Stochastic Dominance, Poverty and the Treatment Effect Curve," *Economics Bulletin*, 30(1), 365–373.

# Appendices

## 2.A Sampling procedure

**Table 2.A.1:** Sampling process.

	Original number of children (a)	matched children number (b)	% of (a)	1997 data available number	% of (b)
C	2247	1871	83	1871	100
T	2615	2200	84	1128	51
Total	4862	4071	84	2999	73

Note: the acronyms refer to samples : C = Control sample; T = Treatment sample.

Comparing the sample sizes in the column “1997 data available” with those in table 1 in the main text, one observes that 12 (3) observations dropped out in the final control (treatment) sample. This is due to missing observations on circumstance characteristics.

## 2.B Results of the logistic regression

Our specification for the logistic regression is close to the specification used for propensity score matching by [Behrman, Parker, and Todd \(2009b\)](#) and [Behrman and Parker \(2010\)](#). The dependent variable equals 1 if the observation comes from the control sample and 0 otherwise. Explanatory variables are based on pre-program characteristics of the treatment sample and recalled 1997 characteristics of the control sample. We have five kinds of explanatory variables.

- (1) Household characteristics: the ages of the household head and spouse (in years), sex of household head, whether the household head and spouse speak an indigenous language, whether they completed primary education, whether they worked, and composition of the household (number of children, women and men of different ages).
- (2) Dwelling conditions of the household: number of rooms in the house and a list of dummy variables indicating the presence of electrical light, running water on the property, running water in the house (which implies of course presence of running water on the property), a dirt

floor and whether the roof and wall were of poor quality.

(3) Asset information: dummy variables indicating whether the family owned animals or land and whether the family possessed a blender, fridge, fan, gas stove, gas heater, radio, hifi, TV, video, washing machine, car or truck.

(4) State of residence: a list of dummy variables indicating the state where the family lived. The reference state (all state of residence dummies equal to zero) is Veracruz.

(5) Following [Behrman, Parker, and Todd \(2009b\)](#) and [Behrman and Parker \(2010\)](#), we include dummy variables for missing characteristics, provided they could be meaningfully estimated. The variable "Miss Asset" takes the value of one if any of the assets listed in the table between "Animals" and "Truck" is missing.

Table 2.B.1 gives the estimated coefficients.

Table 2.B.1: Logistic regression results.

Variable	Coef.	St.Er.	z	Variable	Coef.	St.Er.	z
Age Hh head	-0.013	0.007	-1.96	Blender	-0.169	0.132	-1.27
Age spouse	-0.012	0.007	-0.61	Fridge	0.054	0.200	0.27
Sex Hh head	-2.197	0.351	-6.25	Fan	0.142	0.120	0.71
IndigHhHead	-0.718	0.272	-2.64	Gas stove	0.377	0.145	2.60
IndigSpouse	0.249	0.278	0.90	Gas heater	0.709	0.360	1.97
EducHhHead	-0.229	0.114	-2.01	Radio	-0.600	0.100	-5.96
EducSpouse	-0.386	0.116	-3.32	Hifi	-0.361	0.251	-1.44
Work Hh head	1.124	0.262	4.29	Tv	-0.635	0.118	-5.53
Work spouse	0.623	0.161	3.86	Video	0.498	0.345	1.44
# Children 0-5	-0.090	0.048	-1.89	Wash machine	-0.35	0.330	-0.11
# Children 6-12	-0.211	0.042	-5.06	Car	1.229	0.465	2.64
# Children 13-15	-0.160	0.084	-1.91	Truck	0.243	0.282	0.86
# Children 16-20	-0.016	0.073	-0.22	Guerrero	-0.548	0.190	-2.88
# Women 20-39	-0.014	0.119	-0.12	Hidalgo	-0.937	0.209	-4.48
# Women 40-59	0.040	0.155	0.26	Michoacan	-0.582	0.176	-3.30
# Women 60+	0.040	0.185	0.22	Puebla	-1.097	0.150	-7.33
# Men 20-39	-0.162	0.106	-1.54	Queretaro	0.119	0.219	0.54
# Men 40-59	0.366	0.161	2.28	San Luis	-0.462	0.153	-3.02
# Men 60+	0.698	0.234	2.99	Miss Age Sp	-4.297	0.713	-6.03
# Rooms	-0.006	0.010	-0.58	Miss Indig HH	0.799	1.959	0.41
Electrical light	0.036	0.115	0.32	Miss Indig Sp	-2.102	1.894	-1.11
Running water land	0.879	0.115	7.67	Miss Work HH	3.461	1.871	1.85
Running water house	-0.435	0.208	-2.10	Miss Work Sp	3.817	1.844	2.07
Dirtfloor	0.096	0.118	0.81	Miss Water land	0.871	1.640	0.53
Poor quality roof	-0.026	0.108	-0.24	Miss Water house	0.699	0.827	0.84
Poor quality wall	-0.483	0.126	-3.82	Miss Assets	-4.121	2.398	-1.72
Animals	-0.168	0.113	-1.48	Constant	3.860	0.4223	9.13
Land	-0.545	0.105	-5.17				
Number of Obs	2741			Pseudo R2	0.198		
LR Chi2 (54)	730.0			Log Likelihood	-1478.75		
Prob>Chi2	0.000						

## 2.C Matching estimator and construction of the corresponding distribution function.

**Table 2.C.1:** Propensity score matching: common support and number of observations in the common support.

	Common support	Control #	Treatment #	Bandwidth
IP	[0.158,0.957]	155	193	0.074
IL	[0.106,0.868]	228	260	0.074
NP	[0.063,0.949]	668	318	0.071
NL	[0.017,0.952]	586	318	0.071
Total		1637	1089	

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

STEP 1 : Propensity score matching.

The estimated logistic regressions allow us to compute for each observation the propensity score  $P_i$ , the probability that the observation is in the control sample, given its pre-program characteristics  $x_i$ . Figure 2.C.1 depicts the estimated propensity scores. As we matched for each of the 4 combinations of race and parental level of education the treatment into the control sample, we determined the common support for each of these four comparisons as the overlap of the supports of the control and treatment sample. Table 2.C.1 gives the common support and the number of observations in the common support for each of the types.

We tested the balancing property score using Stata. The optimal number of blocks was 11 and we had 54 explanatory variables, resulting in 594 test. In 14 cases the balancing property was rejected. As an additional test, we rerun the logistic equation from table 2.B.1 using the weighted sample. Only four coefficients out of 54 turned out to be significant. These results are quite encouraging.

STEP 2: Construction of the cumulative distribution function.

Let  $I_1$  denote the set of individuals in the treatment sample,  $I_0$  the set of individuals in the control sample and  $S_P$  the region of common support. The number  $n_0$  gives the number of individuals in the set  $I_0 \cap S_P$ . The outcome of individual  $j$  in the control sample is  $Y_{0j}$  and the outcome of individual  $i$  in the treatment sample is  $Y_{1i}$ . Let  $D = 1$  for program participants and  $D = 0$  for those who don’t participate in the program.

The purpose is to match each individual in the control sample with a weighted average of individuals in the treatment sample. The usual estimator of the average treatment effect

then becomes

$$T = \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} [E(Y_{1j} \mid D = 1, P_j) - Y_{0j}],$$

$$\text{with } E(Y_{1j} \mid D = 1, P_j) = \sum_{i \in I_1} W(i, j) Y_{1i}.$$

The construct  $E(Y_{1j} \mid D = 1, P_j)$  is the outcome of the hypothetical individual matched to individual  $j$ . The average treatment effect can be written as

$$T = \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} \sum_{i \in I_1} W(i, j) Y_{1i} - \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} Y_{0j}.$$

The first term is the average of the matched observations, which attaches to each of the original observations  $Y_{1i}$  a weight

$$\omega_i = \frac{1}{n_0} \sum_{j \in I_0 \cap S_P} W(i, j).$$

It is therefore natural (and consistent with the standard model of the estimation of average treatment effects) to use for each observation  $Y_{1i}$  the weight  $\omega_i$  to construct the cumulative distribution function.

There exist many possible ways to determine the weights  $W(i, j)$ . We use a Kernel estimator, such that

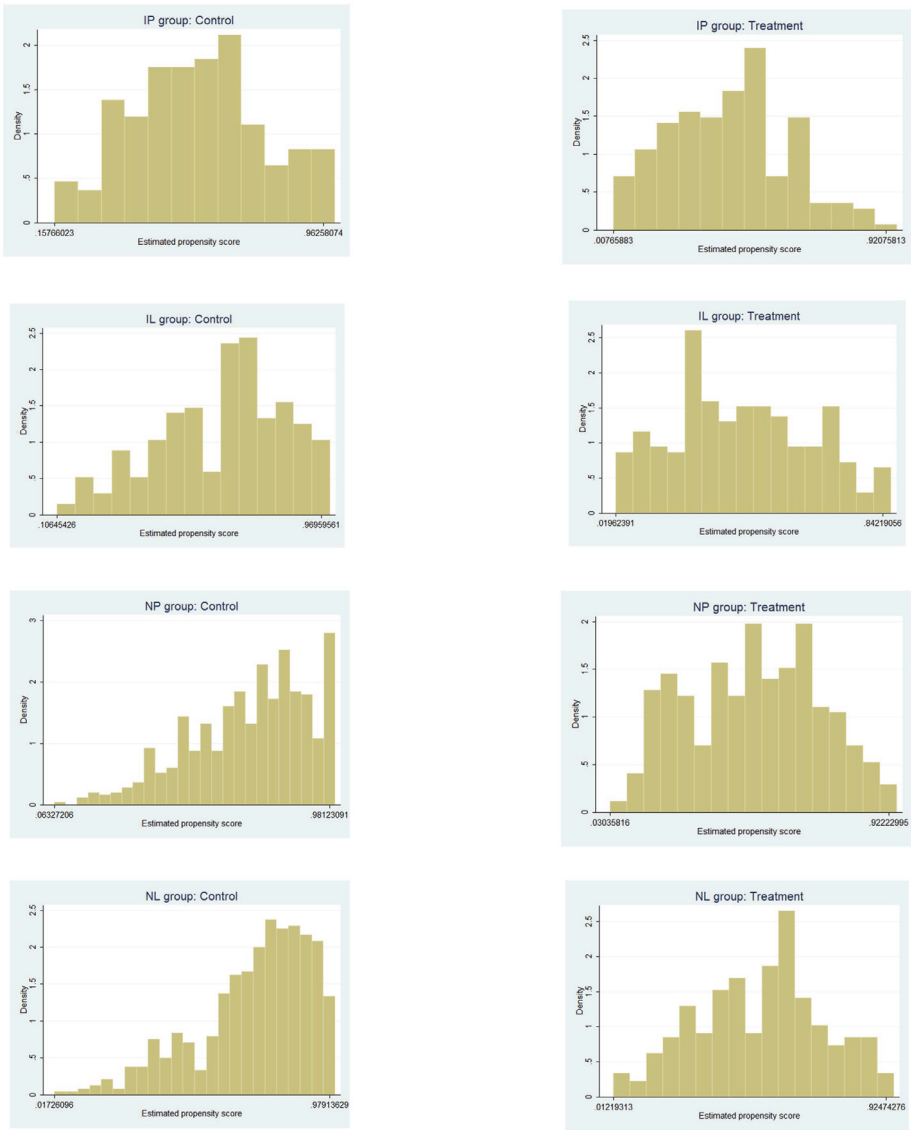
$$W(i, j) = \frac{G\left(\frac{P_i - P_j}{\alpha}\right)}{\sum_{k \in I_1} G\left(\frac{P_k - P_j}{\alpha}\right)},$$

where  $G(\cdot)$  is the Epanechnikov kernel function and  $\alpha$  is a bandwidth parameter. The bandwidth parameter was chosen in an optimal way, using the formula in [Silverman \(1986\)](#), page 45-47:

$$\alpha = 1.06 \min\left(\sigma, \frac{\rho}{1.34}\right),$$

where  $\sigma$  is the standard deviation and  $\rho$  the interquartile range of the distribution of propensity scores. The resulting bandwidths for each of the types are given in the last column of table 2.C.1.

Figure 2.C.1: Estimated propensity scores.





## 2.D Testing stochastic dominance

We explain the approach by focusing on tests for first order stochastic dominance of  $F^T$  over  $F^C$ . Davidson (2009) shows how the approach must be generalized to test for stochastic dominance of arbitrary order.

It is assumed that samples of the control and treatment types that are compared are independent, and their weighted empirical distribution functions  $\hat{F}^C$  and  $\hat{F}^T$  are defined in the usual way. If for the empirical distribution functions  $\hat{F}^C$  and  $\hat{F}^T$ , there exists a  $y \in Y$  such that  $\hat{F}^T(y) \geq \hat{F}^C(y)$ , there is non-dominance in the sample and we do not wish to reject the null.

Davidson and Duclos (2009) restrict the test to a test of the frontier of the null hypothesis against the alternative hypothesis of dominance of  $T$  over  $C$ . The frontier of the null hypothesis is the case where  $\hat{F}^C(y) > \hat{F}^T(y)$  for all  $y \in Y$  except for one point  $y^*$  where  $\hat{F}^C(y^*) = \hat{F}^T(y^*)$ . They show that, for configurations of non-dominance that are not on the frontier, the rejection probabilities of their test are no greater than they are for configurations on the frontier.

For each point in  $R$ , we calculate an unconstrained empirical likelihood ratio statistic and a constrained empirical likelihood ratio statistic, the statistic under the frontier of the null (i.e. imposing the null of non-dominance). The square root of the double difference between these two statistic is the test statistic.<sup>22</sup> Denote this value by  $LR$ . Next, determine the value for which  $LR$  is minimal, as this is the most likely point at which non-dominance cannot be rejected and compute the probabilities  $p_t^X$  associated with each point in sample  $X$  ( $x = C, T$ ) that maximizes the empirical likelihood function subject to  $\hat{F}^C(y^*) = \hat{F}^T(y^*)$ . These probabilities are estimates of the population probabilities under the assumption of non-dominance and are used to set up the following bootstrap data-generating process on the frontier of the null of non-dominance.

We compute 3000 bootstrap samples from the two distributions  $p_t^C$  and  $p_t^T$ , following the original sample design, as suggested by Chen and Duclos (2008). Our samples contain  $C_1^X, \dots, C_c^X, \dots, C_{n^X}^X$  clusters (villages),  $X = C, T$ . Each cluster in the sample contains  $n_c^X$  children ( $c = 1, \dots, n^X$ ). We mimic this sample design as follows. First, define for each cluster

$$\pi_c^X = \frac{\sum_{t \in C_c^X} p_t^X}{\sum_{t \in \bigcup_{c=1 \dots n^X} C_c^X} p_t^X},$$

<sup>22</sup>For first order stochastic dominance, this statistic can be analytically obtained. For second order dominance the statistic has to be numerically determined using the Newton method to solve a set of non-linear equations -see Davidson (2009).

which gives the probability that an observation is drawn from cluster  $c$ . Now, draw the identity of the first cluster from the  $n^X$  clusters, such that each cluster has a probability  $\pi_c^X$  of being drawn. This gives, say cluster  $k$ . Next, draw  $n_1^X$  observations from cluster  $k$  with replacement, where each observation has a probability  $p_t^k / \sum_{t \in C_k^X} p_t^X$  of being drawn. Do the same for all the other  $n^X - 1$  clusters. This gives the first bootstrap sample. Repeat the procedure 3000 times. For each bootstrap sample, we calculate the minimal  $LR$  statistic to get an idea of the distribution of the minimal  $LR$  under the frontier of the null hypothesis. The  $p$ -value of the sample statistic is then the fraction of bootstrap-statistics greater than the sample statistic.

When there is dominance in the sample, we report the results by giving the longest interval  $[\hat{r}^-, \hat{r}^+]$  for which the hypothesis

$$\max_{z \in [\hat{r}^-, \hat{r}^+]} \left( F^T(z) - F^C(z) \right) \geq 0,$$

can be rejected. For a given level of significance  $\alpha$ ,  $\hat{r}^-$  ( $\hat{r}^+$ ) is the smallest (greatest) value of  $r^-$  ( $r^+$ ) for which the hypothesis

$$\max_{z \in [r^-, r^+]} \left( F^T(z) - F^C(z) \right) \geq 0$$

can be rejected at level  $\alpha$ . The larger is this interval, given  $\alpha$ , the more powerful our rejection of non-dominance. We ignore the stochastic nature of the sampling weights.

## 2.E Roemer's identification axiom and matching estimator (weighted treatment distribution)

### (1) The standard Roemer model and its assumptions

In the standard model health,  $h$ , is determined only by parental circumstances,  $c$ , and a scalar representing parental responsibility,  $p$ :

$$h(c, p).$$

Define, for each type  $\hat{h}^i$  as the level of health such that a fraction  $R$  of type  $i$  has a health not better than  $R$ :

$$\int_P I \left( h(c_i, p) \leq \hat{h}^i \right) f_p^i(p) dp = R, \quad (2.1)$$

where  $I(\cdot)$  is the indicator function. The first assumption typically made to derive Roemer's identification axiom is

**A1:**  $h(c, p)$  is strictly increasing in  $p$ .

As a result of this assumption, there exists for each type a value  $p^i$  such that

$$I\left(h(c_i, p) \leq \hat{h}^i\right) = 1 \Leftrightarrow p \leq p^i,$$

and we get from (2.1),

$$R = \int_{\underline{p}}^{p^i} f_p^i(p) dp.$$

Imposing the second assumption,

**A2:** For all  $i$ ,  $f_p^i(p) = f_p(p)$ ,

which says that responsibility is distributed independently from circumstances, we get

**RIA:** For all  $i$ ,  $p^i = p^*$ ,

which is Roemer's identification axiom: those that are at the same percentile in the distribution of health within their type, have the same responsibility.

(2) Weighted treatment observations and a variant of RIA

Suppose children's health is influenced by parental circumstances,  $c$ , pre-program characteristics,  $x$ , and a scalar representing parental responsibility,  $p$ :

$$h(c, x, p).$$

Define for the treatment sample after weighting the observations the value  $\hat{h}^T$  and for the control sample the value  $\hat{h}^C$  such that the same fraction in both samples has a health smaller than or equal to these critical values.

$$\begin{aligned} \int_P \int_X I\left(h(c^T, x, p) \leq \hat{h}^T\right) \tilde{f}_{x,p}^T(x, p) dx dp = \\ \int_P \int_X I\left(h(c^C, x, p) \leq \hat{h}^C\right) f_{x,p}^C(x, p) dx dp, \end{aligned} \quad (2.2)$$

where

$$\tilde{f}_{x,p}^T(x, p) = f_{p|x}^T(p|x) f_x^C(x),$$

the joint distribution of  $x$  and  $p$  after weighting the observations in the treatment sample, which ensures that the marginal distribution of  $x$  is the same in the control and treatment sample. A first assumption that can be made is

$$\mathbf{A3}: f_{p|x}^T(p|x) = f_{p|x}^C(p|x).$$

This says that the distribution of responsibility conditional on  $x$  is the same in the treatment and control group. It implies that

$$\hat{f}_{x,p}^T(x, p) = f_{x,p}^C(x, p). \quad (2.3)$$

As a result, (2.2) reduces to

$$\int_P \int_X \left[ I(h(c^T, x, p) \leq \hat{h}^T) - I(h(c^C, x, p) \leq \hat{h}^C) \right] f_{x,p}^C(x, p) dx dp = 0. \quad (2.4)$$

A second assumption that can be made is that the function  $h(c, x, p)$  is additively separable between  $c$  and  $(x, p)$ .

$$\mathbf{A4}: \text{There exist functions } v(c) \text{ and } w(x, p) \text{ such that } h(c, x, p) = v(c) + w(x, p).$$

This allows us to write (2.4) as

$$\int_P \int_X \left[ I(w(x, p) \leq \hat{h}^T - v(c^T)) - I(w(x, p) \leq \hat{h}^C - v(c^C)) \right] f_{x,p}^C(x, p) dx dp = 0.$$

As this equation must hold for arbitrary distribution functions  $f_{x,p}^C(x, p)$ , it follows that

$$\hat{h}^T - v(c^T) = \hat{h}^C - v(c^C).$$

As a result,

$$\begin{aligned} h(c^T, x, p) = \hat{h}^T &\Leftrightarrow v(c^T) + w(x, p) = \hat{h}^T \Leftrightarrow w(x, p) = \hat{h}^T - v(c^T) \\ &\Leftrightarrow w(x, p) = \hat{h}^C - v(c^C) \Leftrightarrow h(c^C, x, p) = \hat{h}^C. \end{aligned}$$

Now consider the expected value of  $p$  in the weighted treated and control sample, given that health is at the same percentile.

$$E(p|h = \hat{h}^T) = \frac{1}{\hat{f}_h^T(h)} \int_P p \int_X I(h(c^T, x, p) = \hat{h}^T) \hat{f}_{p,x}^T(p, x) dx dp, \quad (2.5)$$

$$E(p|h = \hat{h}^C) = \frac{1}{f_h^C(h)} \int_P p \int_X I(h(c^C, x, p) = \hat{h}^C) f_{p,x}^C(p, x) dx dp. \quad (2.6)$$

We have shown that weighting the treatment sample and A3 implies (2.3) and that A3 together with A4 imply  $h(c^T, x, p) = \hat{h}^T \Leftrightarrow h(c^C, x, p) = \hat{h}^C$ , such that the expressions behind the first integral sign in (2.5) and (2.6) are equal. What needs to be shown is that the marginal distributions  $\hat{f}_h^T(h)$  and  $f_h^C(h)$  are equal. This follows directly from the previous reasoning,

upon observing that

$$\begin{aligned}\widehat{f}_h^T(h) &= \int_P \int_X I(h(c^T, x, p) = \widehat{h}^T) \widehat{f}_{p,x}^T(p, x) dx dp \quad \text{and} \\ f_h^C(h) &= \int_P \int_X I(h(c^C, x, p) = \widehat{h}^C) f_{p,x}^C(p, x) dx dp.\end{aligned}$$

Conclusion: if both assumptions A3 and A4 hold true, then the weighting procedure guarantees that those that are at the same percentile in the distribution of health in the weighted treatment and control sample have the same expected value for responsibility.

## 2.F Treatment and control effects in matched samples

**Table 2.F.1:** Health outcomes of 2-6 year old children in 2003.

(a) Control sample							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.24	12.0	0.32	-1.47	0.24	0.58	0.17
IP	0.36	11.5	0.46	-1.91	0.23	0.54	0.19
IL	0.30	11.9	0.63	-2.36	0.30	0.63	0.13
NP	0.18	12.2	0.19	-1.12	0.21	0.57	0.18
NL	0.24	12.0	0.32	-1.47	0.26	0.58	0.17
(a) Treatment sample							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.20	12.1	0.32	-1.47	0.19	0.67	0.11
IP	0.19	12.0	0.30	-1.52	0.14	0.66	0.12
IL	0.25	11.7	0.45	-1.86	0.18	0.71	0.07
NP	0.10	12.4	0.24	-1.10	0.25	0.68	0.09
NL	0.25	12.3	0.30	-1.41	0.21	0.64	0.15

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

As expected since we match the treatment sample to the control samples, the characteristics of the matched control sample are very similar to those of the original control sample in table 2.2. The differences between the matched and original treatment sample are larger.

## 2.G Sensitivity Analysis

In addition to the results discussed at length in the main part of the paper, we present three sensitivity analyses by modifying inclusion criteria to the program and by modifying the definition of parental education. In the analysis presented in sections 2.3 and 2.4, our base case, two conditions were necessary for inclusion to the treatment group: (i) the household should be part of a treatment community (communities where the program was operating) and (ii) information on monetary transfers received by the household should be available. Children's types were defined on the basis of indigenous origin and whether at least one parent completed primary education or not.

The analysis in appendix 2.G.1 that follows incorporates *all* children living in treatment communities independently of whether information on transfers received by the household were available or not. As a result, the treatment sample for this analysis contains 219 additional observations, as can be seen upon simple comparison of tables 2.1 (in the main text) and 2.G.1.1 (in appendix 2.G.1).

Comparing the results in tables 2.3 and 2.G.1.4, it is striking that all estimated program effects have the same sign. Most significantly estimated effects in table 2.1 also turn up significant in table 2.G.1.4 and the other way around, with few exceptions. All significantly estimated effects in both tables 2.3 and 2.G.1.4 are in favor of the program. Looking at the stochastic dominance results in figure 2.G.1.1 we find very similar arrangements as in the base case in figure 2.1. Indigenous children seem to benefit most from receiving Oportunidades, although the effect now is somewhat weaker for indigenous children without parental primary education background (IL group) and stronger for indigenous children with parental primary education background (IP group). Except for the negative effect observed on standardized BMI for non-indigenous children with parental primary education, the effect on non-indigenous is similar to the base case. Overall, the effects are very close to the effects in the base case and very much in favor of the program.

The analysis in appendix 2.G.2 changes the definition of type. Although the contribution of parental education to child health is generally recognized, education effects of both parents separately are still disputed (Breierova and Duflo (2003), Aslam and Kingdon (2010)). In particular, it has been suggested that education of the mother could have a major influence on child well-being (Desai and Alva (1998)). Based on this hypothesis, appendix 2.G.2 defines types on the basis of indigenous background (as before) and on whether the *mother* has completed primary education or not. Table 2.G.2.1 shows that this diminishes the sizes of both control and treatment samples compared to the base case (table 2.1). This is due to the

fact that in the base case, some observations for which mother's educational level was not but father's educational level was observed, could be classified as having at least one parent that completed primary education.

The comparison of tables 2.3 and 2.G.2.4 reveals that all estimated program effects have the same sign (except the effect on the fraction of anemic children for the NL group, which changes from being marginally positive to -0.02). Most significantly estimated effects in table 2.1 also turn up significant in table 2.G.2.4 and the other way around, with only few exceptions. All significantly estimated effects in table 2.G.2.4 are in favor of the program. The stochastic dominance tests in figure 2.G.2.1 show the same pattern as in figure 2.1. The most noteworthy difference is that the positive effects on hemoglobin concentration and standardized height of the IL group become less pronounced. From this sensitivity analysis, we conclude again that, overall, the effects are very close to the effects in the base case and very much in favor of the program.

The analysis in appendix 2.G.3 compares the effect of Oportunidades between the *immediate and delayed treatment* groups. As mentioned in section 2.3.2, the original sample design followed a random procedure to allocate the treatment to two comparable groups. One group received the program immediately (original treatment) while the other was phased-out 18 months in order to operate as control (delay treatment). Lack of information on the amount of transfers for the original treatment motivated the use of the latter for the main analysis. Here we aim at assessing the effect of having been exposed longer to the program, by comparing the health outcomes of children in the original and delayed treatment. The main advantage is the randomization of households over these two groups. In order to make the comparison meaningful, we limit the sample to children that were born after April 1998 (when the original treatment started) and before October 1999 (when delayed treatment started). As can be seen in table 2.G.3.1, this decreases the number of observations that can be used drastically.

The logistic regression in table 2.G.3.5 reports much fewer significant coefficients than the regressions in tables 2.B.1, 2.G.1.5 and 2.G.2.5. This is due to the much better randomization of households between initial and delayed treatment and the smaller sample size. Table 2.G.3.4 shows the limitation of the exercise: it shows only one positive treatment effect: for indigenous children with a mother that completed primary education, the fraction of children reporting zero sick days increased by 18 percent. Also the stochastic dominance tests find fewer significant effects, especially on hemoglobin concentration and days sick. The reason for this is probably that both the children in the initial and delayed treatment samples received the program during the three years preceding the collection of the health data in 2003, and these

**Table 2.G.1.1:** Composition of the samples (delay vs control).

	Control sample		Treatment sample	
	#	%	#	%
All	1859	100	1344	100
IP	173	9.3	227	16.9
IL	241	13.0	329	24.5
NP	824	44.3	395	29.4
NL	621	33.4	393	29.2

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

two health indicators are more influenced by what happens during the period immediately before they are measured. What is quite remarkable, however is the substantial impact on standardized height of longer exposure to the program, exposure in womb and during the first months of life. Here we do find significant effects for children from indigenous origin without a parent that completed primary education and for non-indigenous children with a parent that completed primary education. Hence, we conclude that program exposure at a very young age can have significant positive effects on standardized health three years later.

**2.G.1 Entire delayed treatment group versus Control**



**Table 2.G.1.2:** Health outcomes of 2-6 year old children in 2003 (delay vs control).

(a) Control sample							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.24	12.0	0.32	-1.46	0.24	0.58	0.17
IP	0.36	11.6	0.50	-1.99	0.23	0.57	0.19
IL	0.30	11.9	0.64	-2.40	0.30	0.64	0.13
NP	0.18	12.2	0.20	-1.13	0.22	0.56	0.18
NL	0.25	12.0	0.32	-1.47	0.25	0.58	0.18
(b) Treatment sample							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.23	12.1	0.33	-1.53	0.20	0.66	0.12
IP	0.27	12.0	0.36	-1.70	0.14	0.62	0.13
IL	0.29	11.7	0.45	-1.87	0.18	0.70	0.10
NP	0.14	12.5	0.24	-1.17	0.25	0.67	0.11
NL	0.26	12.2	0.30	-1.47	0.22	0.64	0.14

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

**Table 2.G.1.3:** Health outcomes of 2-6 year old children in 2003 (delay vs control): Matched samples.

(a) Control sample							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.24	12.0	0.32	-1.46	0.24	0.58	0.18
IP	0.37	11.6	0.49	-1.99	0.23	0.57	0.19
IL	0.30	11.9	0.64	-2.37	0.30	0.64	0.13
NP	0.18	12.2	0.19	-1.13	0.22	0.56	0.18
NL	0.25	12.0	0.32	-1.46	0.25	0.58	0.18
(a) Treatment sample							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.22	12.0	0.33	-1.51	0.19	0.67	0.11
IP	0.20	12.0	0.29	-1.54	0.13	0.67	0.10
IL	0.29	11.7	0.48	-1.90	0.18	0.72	0.09
NP	0.11	12.4	0.24	-1.12	0.25	0.62	0.11
NL	0.28	12.2	0.32	-1.45	0.20	0.65	0.15

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

**Table 2.G.1.4:** Difference between control and treatment in fraction of anemic, stunted at risk of being overweight and days sick, weighted samples (delay vs control).

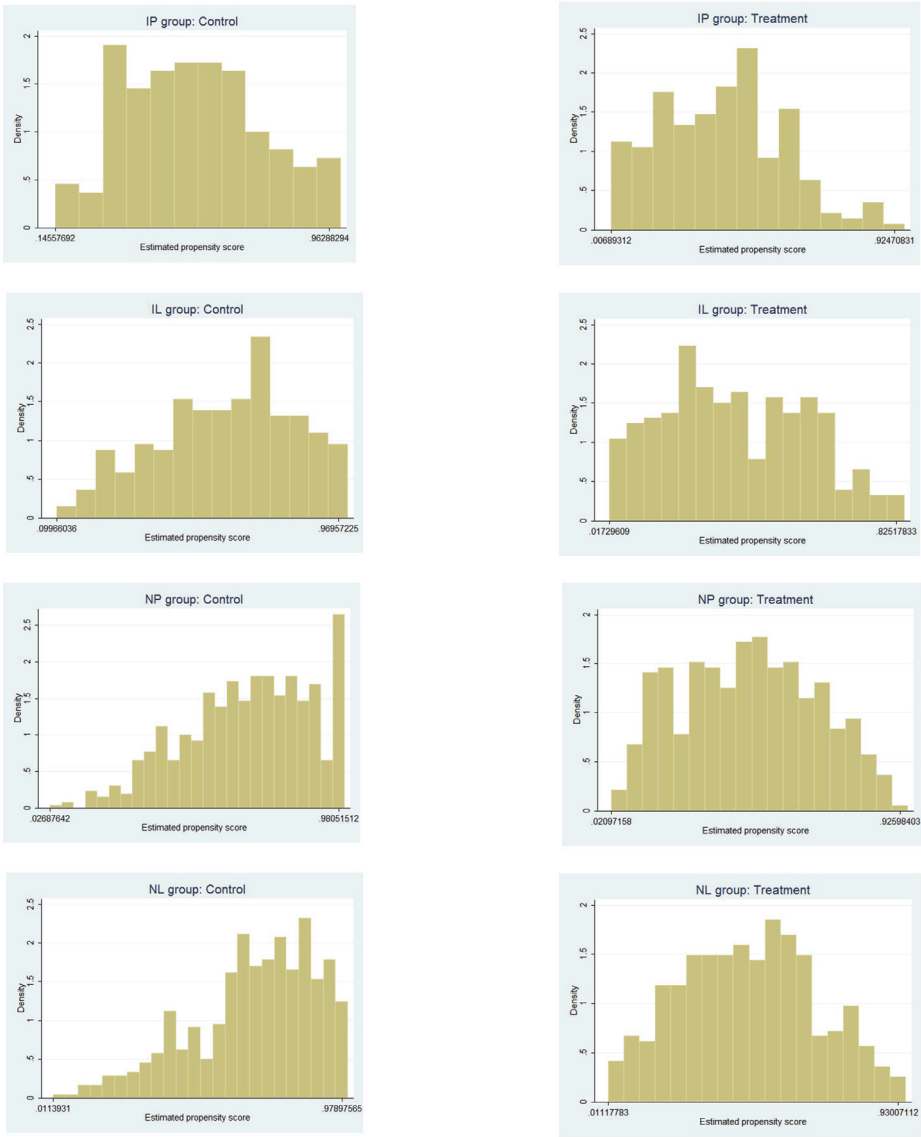
	Anemic	Stunted	Risk Overweight	0 Days Sick	> 3 Days Sick
All	-0.02	0.01	-0.05*	0.09**	-0.06**
IP	-0.16**	-0.21**	-0.10	0.10	-0.08*
IL	-0.01	-0.16*	-0.12**	0.08	-0.05
NP	-0.07*	0.05	0.04	0.06	-0.07**
NL	0.03	-0.00	-0.05*	0.07	-0.03

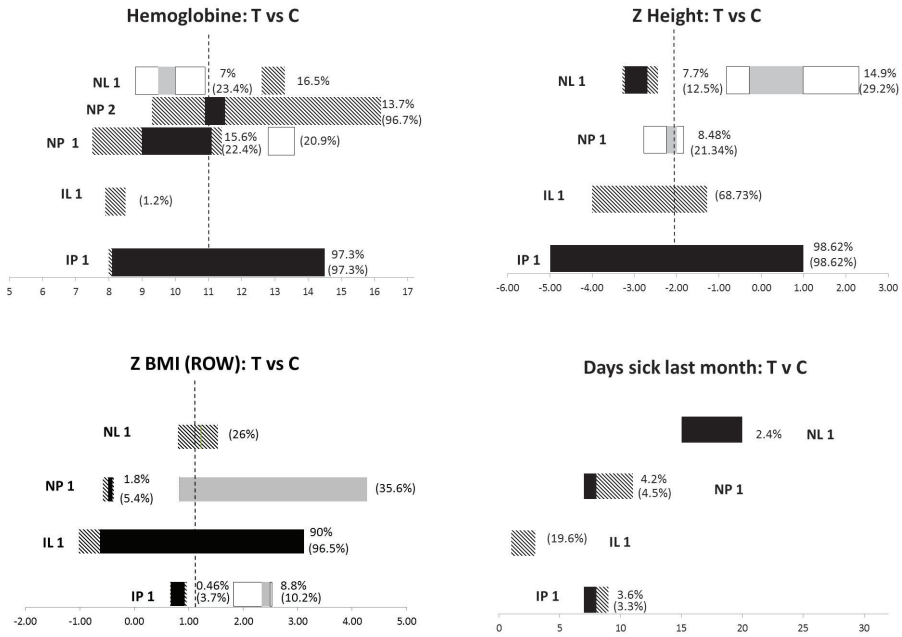
Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education. One (two) “\*\*” indicates that the effect is statistically significant from zero at the ten (five) percent level. Standard errors corrected for clustering at locality level.

Table 2.G.1.5: Logistic regression results (delay vs control).

Variable	Coef.	St.Er.	z	Variable	Coef.	St.Er.	z
Age Hh head	-0.016	0.006	-2.44	Blender	-0.186	0.126	-1.48
Age spouse	-0.014	0.007	-2.12	Fridge	-0.032	0.184	-0.17
Sex Hh head	-2.32	0.335	-6.94	Fan	0.167	0.185	0.90
IndigHhHead	-0.600	0.255	-2.34	Gas stove	0.276	0.136	2.03
IndigSpouse	0.109	0.260	0.42	Gas heater	0.707	0.310	2.28
EducHhHead	-0.234	0.110	-2.13	Radio	-0.546	0.096	-5.67
EducSpouse	-0.487	0.110	-4.39	Hifi	-0.360	0.230	-1.56
Work Hh head	1.024	0.244	4.20	Tv	-0.646	0.114	-5.67
Work spouse	0.490	0.147	3.32	Video	0.412	0.293	1.41
# Children 0-5	-0.077	0.045	-1.71	Wash machine	-0.200	0.290	-0.69
# Children 6-12	-0.182	0.040	-4.58	Car	0.877	0.358	2.45
# Children 13-15	-0.156	0.080	-1.98	Truck	-0.243	0.236	-1.03
# Children 16-20	-0.085	0.067	-1.27	Guerrero	-0.841	0.171	-4.91
# Women 20-39	-0.126	0.105	-1.20	Hidalgo	-0.863	0.196	-4.40
# Women 40-59	-0.083	0.142	-0.59	Michoacan	-0.422	0.167	-2.52
# Women 60+	0.016	0.171	0.09	Puebla	-1.061	0.142	-7.43
# Men 20-39	-0.200	0.096	-2.07	Queretaro	0.290	0.207	1.40
# Men 40-59	0.420	0.152	2.76	San Luis	-0.460	0.144	-3.18
# Men 60+	0.757	0.221	3.42	Miss Age Sp	-4.65	0.712	-6.54
# Rooms	-0.006	0.010	-0.68	Miss Indig HH	0.911	2.08	0.44
Electrical light	0.083	0.110	0.75	Miss Indig Sp	-2.30	2.030	-1.13
Running water land	0.812	0.108	7.51	Miss Work HH	3.65	2.004	1.82
Running water house	-0.350	0.191	-1.84	Miss Work Sp	3.990	1.984	2.01
Dirtfloor	0.060	0.112	0.54	Miss Water land	1.090	1.601	0.68
Poor quality roof	-0.002	0.104	-0.02	Miss Water house	0.354	0.643	0.55
Poor quality wall	-0.380	0.123	-3.09	Miss Assets	-3.912	2.070	-1.89
Animals	-0.191	0.107	-1.78	Constant	4.232	0.406	10.42
Land	-0.505	0.100	-5.06				
Number of Obs	2959						
LR Chi2 (54)	815.6			Pseudo R2	0.201		
Prob>Chi2	0.000			Log Likelihood	-1624.23		

Figure 2.G.1.1: Estimated propensity scores (delay vs control).



**Figure 2.G.1.2:** Stochastic dominance results (delay vs control).**Table 2.G.1.6:** Propensity score matching: common support and number of observations in the common support (delay vs control).

	Common support	Control #	Treatment #	Band-width
IP	[0.145,0.959]	148	209	0.072
IL	[0.099,0.850]	191	312	0.069
NP	[0.027,0.950]	596	392	0.068
NL	[0.011,0.950]	551	390	0.069
Total		1486	1303	

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

2.G.2 Mother’s education as circumstance criterion

Table 2.G.2.1: Composition of the samples  
(Mother’s education case).

	Control sample		Treatment sample	
	#	%	#	%
All	1808	100	1079	100
IP	121	6.7	150	13.9
IL	278	15.4	310	28.7
NP	680	37.6	255	23.6
NL	729	40.3	364	33.8

Note: the acronyms refer to types : IP = Indigenous, Mother’s primary education; IL = Indigenous, Mother’s lower education; NP = Non-indigenous, Mother’s primary education; NL = Non-indigenous, Mother’s lower education.

Table 2.G.2.2: Health outcomes of 2-6 year old children in 2003  
(Mother’s education case).

(a) Control sample							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.24	12.0	0.32	-1.46	0.24	0.58	0.17
IP	0.38	11.9	0.48	-1.88	0.24	0.56	0.18
IL	0.30	11.8	0.63	-2.36	0.28	0.63	0.14
NP	0.17	12.3	0.17	-1.04	0.23	0.57	0.17
NL	0.25	12.0	0.33	-1.52	0.23	0.58	0.18
(b) Treatment sample							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.22	12.1	0.33	-1.58	0.20	0.67	0.12
IP	0.31	11.9	0.34	-1.64	0.12	0.66	0.14
IL	0.27	11.8	0.43	-1.83	0.17	0.72	0.10
NP	0.10	12.6	0.23	-1.25	0.22	0.68	0.11
NL	0.23	12.2	0.32	-1.58	0.24	0.64	0.14

Note: the acronyms refer to types : IP = Indigenous, Mother’s primary education; IL = Indigenous, Mother’s lower education; NP = Non-indigenous, Mother’s primary education; NL = Non-indigenous, Mother’s lower education.

2.G.3 Original versus Delay treatment

**Table 2.G.2.3:** Health outcomes of 2-6 year old children in 2003  
(Mother's education case): Matched samples.

(a) Control sample							
	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.24	12.0	0.32	-1.46	0.24	0.58	0.17
IP	0.38	11.9	0.48	-2.00	0.24	0.56	0.18
IL	0.30	11.8	0.63	-2.36	0.28	0.63	0.14
NP	0.17	12.3	0.17	-1.04	0.23	0.57	0.17
NL	0.25	12.0	0.33	-1.51	0.23	0.58	0.18

(a) Treatment sample							
	Hemoglobin		zheight		zBMI	Days Sick	
	Anemic	Median	Stunted	Median	ROW	0	> 3
All	0.20	12.1	0.32	-1.47	0.19	0.67	0.11
IP	0.25	12.0	0.26	-1.46	0.12	0.66	0.17
IL	0.25	11.9	0.50	-1.97	0.17	0.74	0.07
NP	0.07	12.4	0.21	-1.09	0.23	0.61	0.11
NL	0.23	12.3	0.32	-1.46	0.22	0.66	0.12

Note: the acronyms refer to types : IP = Indigenous, Mother's primary education; IL = Indigenous, Mother's lower education; NP = Non-indigenous, Mother's primary education; NL = Non-indigenous, Mother's lower education.

**Table 2.G.2.4:** Difference between control and treatment in fraction of anemic, stunted at risk of being overweight and days sick, weighted samples (Mother's education case).

	Anemic	Stunted	Risk Overweight	0 Days Sick	> 3 Days Sick
All	-0.04	0.00	-0.05*	0.09**	-0.06**
IP	-0.12	-0.21**	-0.12*	0.10	-0.02
IL	-0.05	-0.13	-0.12**	0.11**	-0.08**
NP	-0.10**	0.05	0.00	0.04	-0.06**
NL	-0.02	-0.00	-0.01	0.08	-0.07**

Note: the acronyms refer to types : IP = Indigenous, Mother's primary education; IL = Indigenous, Mother's lower education; NP = Non-indigenous, Mother's primary education; NL = Non-indigenous, Mother's lower education. One (two) "\*" indicates that the effect is statistically significant from zero at the ten (five) percent level. Standard errors corrected for clustering at locality level.

Table 2.G.2.5: Logistic regression results (Mother’s education case).

Variable	Coef.	St.Er.	z	Variable	Coef.	St.Er.	z
Age Hh head	-0.017	0.007	-2.40	Blender	-0.180	0.134	-1.34
Age spouse	-0.005	0.007	-0.63	Fridge	0.073	0.204	0.36
Sex Hh head	-2.354	0.380	-6.19	Fan	0.125	0.202	0.62
IndigHhHead	-0.691	0.288	-2.40	Gas stove	0.364	0.147	2.48
IndigSpouse	0.213	0.292	0.73	Gas heater	0.609	0.365	1.67
EducHhHead	-0.222	0.115	-1.92	Radio	-0.590	0.101	-5.79
EducSpouse	-0.398	0.117	-3.39	Hifi	-0.357	0.254	-1.41
Work Hh head	1.199	0.280	4.27	Tv	-0.626	0.120	-5.22
Work spouse	0.575	0.164	3.50	Video	0.572	0.354	1.62
# Children 0-5	-0.090	0.048	-1.84	Wash machine	-0.139	0.337	-0.41
# Children 6-12	-0.218	0.042	-5.13	Car	1.214	0.468	2.59
# Children 13-15	-0.144	0.085	-1.70	Truck	0.262	0.287	0.91
# Children 16-20	-0.218	0.042	-5.13	Guerrero	-0.535	0.191	-2.80
# Women 20-39	-0.008	0.121	-0.70	Hidalgo	-0.864	0.218	-3.96
# Women 40-59	0.036	0.157	0.23	Michoacan	-0.576	0.178	-3.23
# Women 60+	-0.000	0.189	-0.00	Puebla	-1.103	0.151	-7.29
# Men 20-39	-0.178	0.107	-1.65	Queretaro	0.108	0.222	0.49
# Men 40-59	0.036	0.157	0.23	San Luis	-0.441	0.155	-2.83
# Men 60+	0.682	0.241	2.83	Miss Age Sp	-3.85	0.723	-5.33
# Rooms	-0.005	0.010	-0.55	Miss Indig HH	0.649	1.919	0.34
Electrical light	0.059	0.116	0.51	Miss Indig Sp	-2.015	1.850	-1.09
Running water land	0.844	0.116	7.26	Miss Work HH	3.510	1.827	1.92
Running water house	-0.412	0.209	-1.97	Miss Work Sp	3.733	1.795	2.08
Dirtfloor	0.097	0.120	0.81	Miss Water land	0.794	1.627	0.49
Poor quality roof	-0.002	0.109	-0.02	Miss Water house	0.707	0.828	0.85
Poor quality wall	-0.506	0.127	-3.95	Miss Assets	-3.990	2.289	-1.74
Animals	-0.177	0.114	-1.55	Constant	3.900	0.442	8.80
Land	-0.549	0.107	-5.12				
Number of Obs	2635			Pseudo R2	0.190		
LR Chi2 (54)	671.61			Log Likelihood	-1434.70		
Prob>Chi2	0.000						



Figure 2.G.2.1: Estimated propensity scores (Mother's education case).

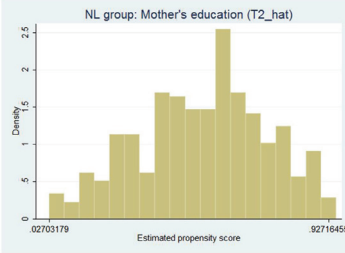
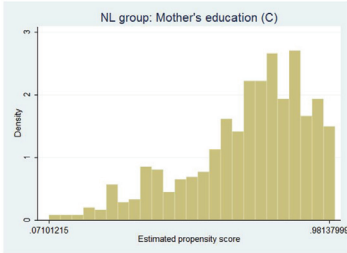
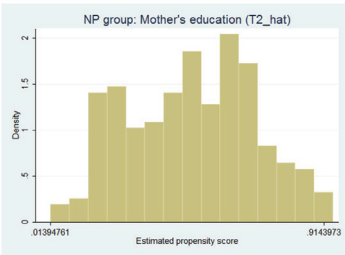
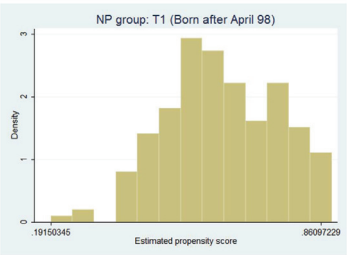
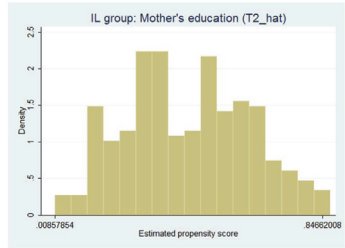
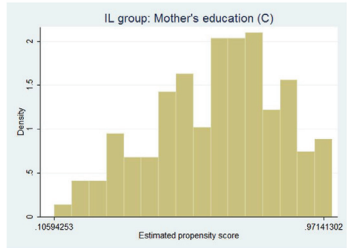
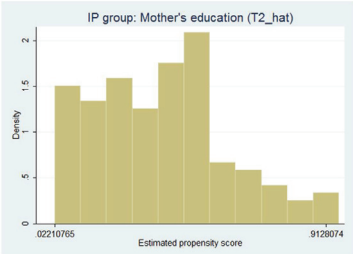
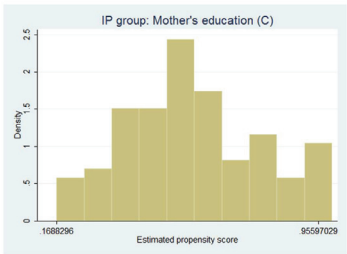


Figure 2.G.2.2: Stochastic dominance results (Mother’s education case).

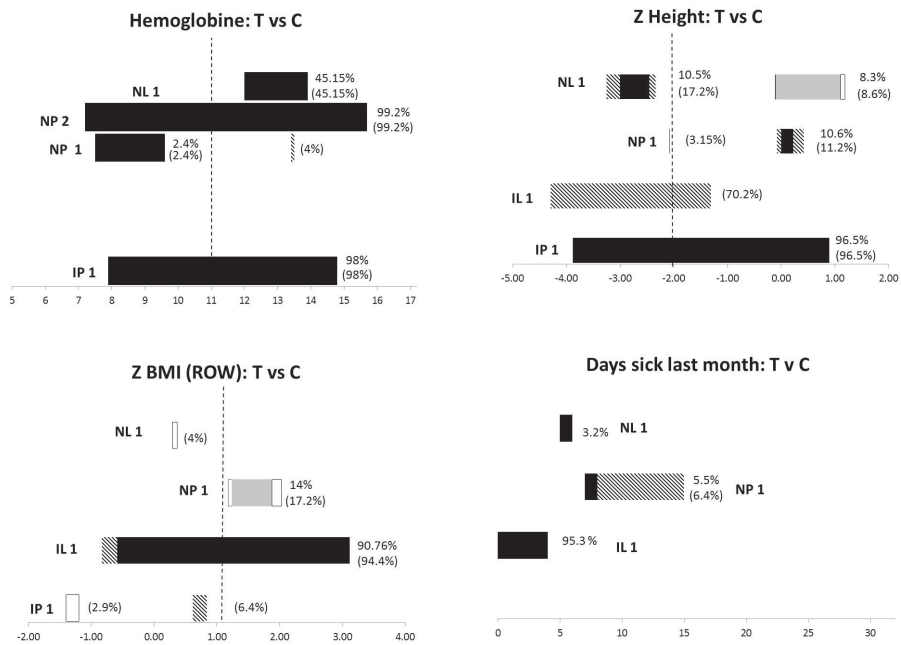


Table 2.G.2.6: Propensity score matching: common support and number of observations in the common support (Mother’s education case).

	Common support	Control #	Treatment #	Band-width
IP	[0.169,0.955]	104	139	0.073
IL	[0.106,0.872]	264	290	0.073
NP	[0.023,0.945]	552	250	0.070
NL	[0.071,0.951]	668	363	0.071
Total		1588	1042	

Note: the acronyms refer to types : IP = Indigenous, Mother’s primary education; IL = Indigenous, Mother’s lower education; NP = Non-indigenous, Mother’s primary education; NL = Non-indigenous, Mother’s lower education.

**Table 2.G.3.1:** Composition of the samples  
(delay vs original treatment).

	Initial treatment		Delayed treatment	
	#	%	#	%
All	730	100	527	100
IP	110	15.1	69	13.2
IL	227	31.1	156	29.5
NP	186	25.5	156	29.5
NL	207	28.3	146	27.7

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

**Table 2.G.3.2:** Health outcomes of 2-6 year old children in 2003  
(delay vs original treatment).

(a) Initial treatment							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.17	12.4	0.35	-1.51	0.17	0.68	0.11
IP	0.27	12.0	0.41	-1.62	0.14	0.72	0.12
IL	0.19	12.4	0.50	-2.00	0.19	0.71	0.11
NP	0.14	12.5	0.24	-1.23	0.18	0.60	0.12
NL	0.13	12.5	0.26	-1.42	0.17	0.71	0.11

(b) Delayed treatment							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.18	12.3	0.33	-1.53	0.17	0.67	0.13
IP	0.31	11.9	0.34	-1.67	0.13	0.61	0.09
IL	0.25	12.0	0.41	-1.81	0.15	0.71	0.09
NP	0.10	12.5	0.23	-1.13	0.21	0.67	0.14
NL	0.14	12.7	0.35	-1.57	0.16	0.64	0.16

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

**Table 2.G.3.3:** Health outcomes of 2-6 year old children in 2003  
(delay vs original treatment): Matched samples.

(a) Initial treatment							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.17	12.4	0.35	-1.51	0.17	0.68	0.11
IP	0.27	12.0	0.40	-1.60	0.14	0.72	0.12
IL	0.19	12.4	0.50	-1.99	0.19	0.71	0.11
NP	0.13	12.5	0.24	-1.23	0.18	0.60	0.12
NL	0.12	12.6	0.26	-1.41	0.17	0.71	0.11

(a) Delayed treatment							
	Hemoglobin		zheight		zBMI ROW	Days Sick	
	Anemic	Median	Stunted	Median		0	> 3
All	0.19	12.4	0.37	-1.65	0.17	0.63	0.14
IP	0.29	12.3	0.42	-1.71	0.16	0.54	0.08
IL	0.25	12.0	0.51	-2.03	0.14	0.71	0.10
NP	0.12	12.5	0.23	-1.14	0.20	0.60	0.22
NL	0.11	12.8	0.34	-1.57	0.17	0.65	0.15

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.

**Table 2.G.3.4:** Difference between initial and delayed treatment in fraction of anemic, stunted at risk of being overweight and days sick, weighted samples (delay vs original treatment).

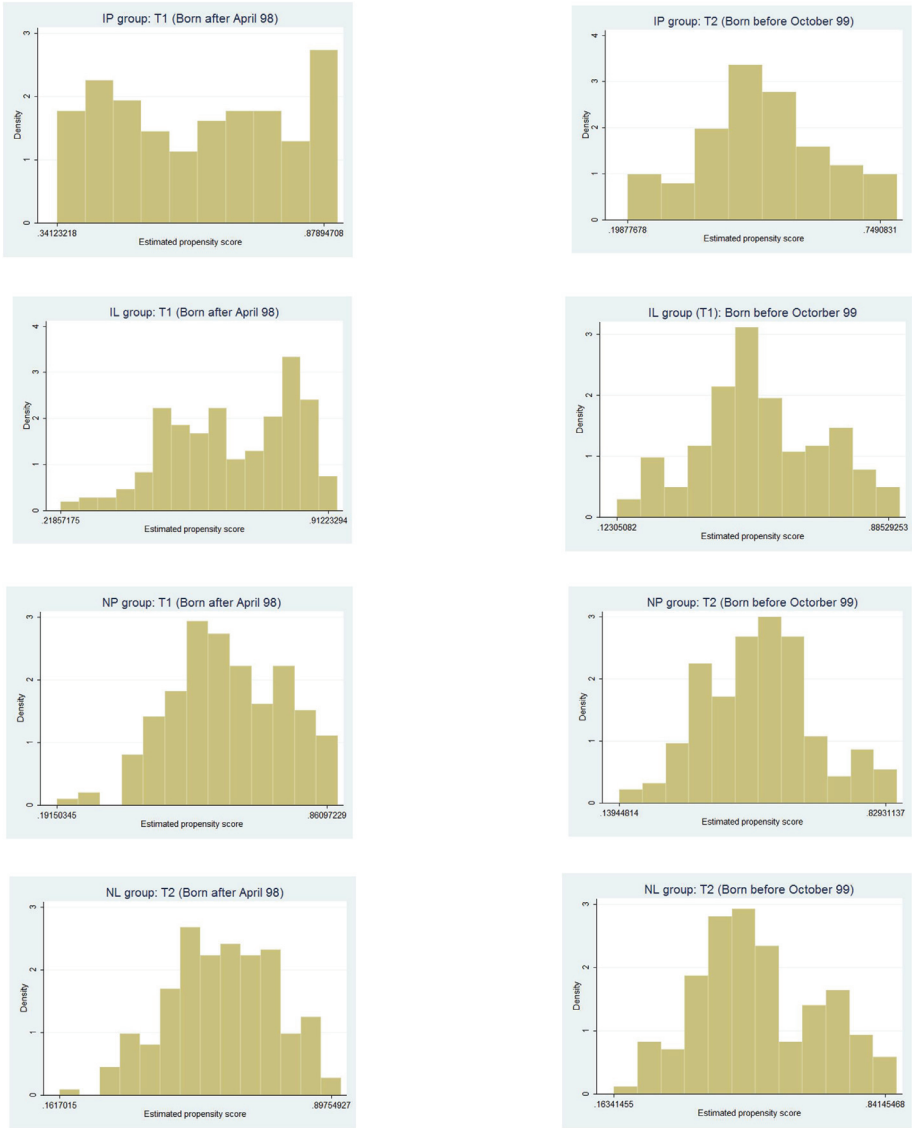
	Anemic	Stunted	Risk Overweight	0 Days Sick	> 3 Days Sick
All	-0.02	-0.02	0.01	0.06	-0.03
IP	-0.02	-0.01	-0.02	0.18**	0.03
IL	-0.06	-0.01	0.04	0.00	0.01
NP	0.02	0.02	-0.02	0.00	-0.10
NL	0.02	-0.07	0.00	0.07	-0.05

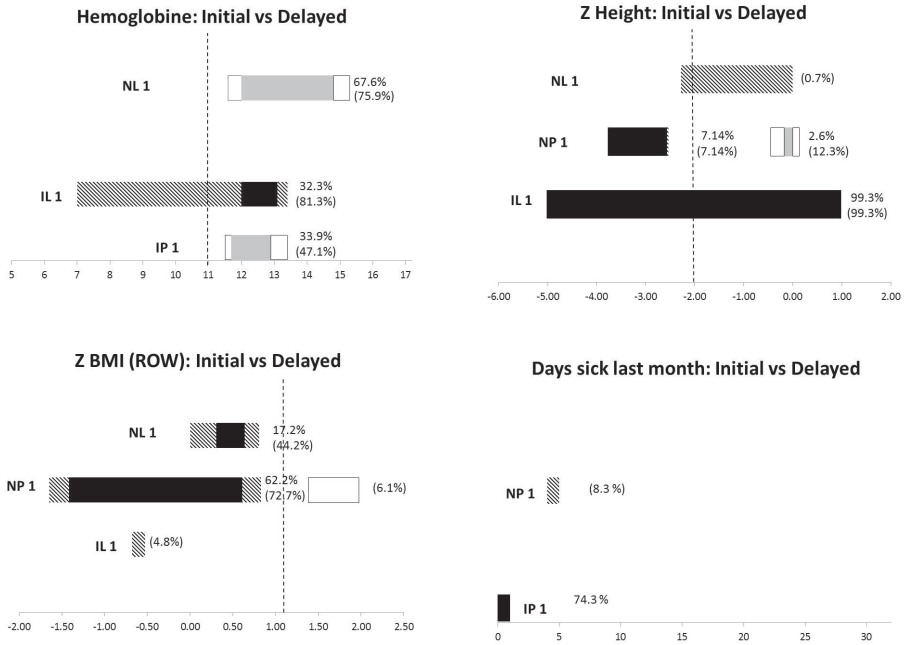
Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education. One (two) \*\* indicates that the effect is statistically significant from zero at the ten (five) percent level. Standard errors corrected for clustering at locality level.

**Table 2.G.3.5:** Logistic regression results (delay vs original treatment).

Variable	Coef.	St.Er.	z	Variable	Coef.	St.Er.	z
Age Hh head	0.015	0.010	1.43	Blender	-0.165	0.180	-0.92
Age spouse	-0.002	0.011	-0.18	Fridge	0.523	0.244	2.15
Sex Hh head	-0.561	0.500	-1.12	Fan	-0.400	0.280	-1.39
IndigHhHead	-0.004	0.314	-0.01	Gas stove	-0.074	0.190	-0.39
IndigSpouse	0.332	0.323	1.03	Gas heater	0.360	0.412	0.87
EducHhHead	-0.085	0.158	-0.54	Radio	-0.040	0.145	-0.28
EducSpouse	-0.136	0.165	-0.82	Hifi	0.044	0.337	0.13
Work Hh head	0.823	0.300	2.74	Tv	-0.110	0.157	-0.70
Work spouse	0.222	0.205	1.08	Video	0.401	0.395	1.02
# Children 0-5	-0.007	0.064	-0.11	Wash machine	-0.236	0.426	-0.55
# Children 6-12	-0.050	0.060	-0.86	Car	0.513	0.589	0.87
# Children 13-15	-0.106	0.119	-0.88	Truck	-0.165	0.296	-0.56
# Children 16-20	-0.025	0.094	-0.27	Guerrero	1.024	0.226	4.52
# Women 20-39	-0.074	0.151	-0.49	Hidalgo	1.596	0.245	6.50
# Women 40-59	-0.347	0.219	-1.58	Michoacan	0.385	0.267	1.44
# Women 60+	-0.208	0.240	-0.87	Puebla	0.630	0.199	3.16
# Men 20-39	-0.099	0.142	-0.70	Queretaro	-0.281	0.349	-0.81
# Men 40-59	-0.150	0.223	-0.65	San Luis	0.506	0.212	2.38
# Men 60+	-0.400	0.315	-1.27	Miss Age Sp	-0.516	1.811	-0.29
# Rooms	-0.018	0.013	-0.42	Miss Indig HH	-0.557	1.792	-0.31
Electrical light	0.066	0.165	0.40	Miss Indig Sp	0.230	2.079	0.11
Running water land	0.356	0.155	2.30	Miss Age Sp	-0.516	1.811	-0.29
Running water house	-0.654	0.292	-2.24	Miss Work Sp	-0.049	1.771	-0.03
Dirtfloor	-0.092	0.161	-0.58	Miss Water land	0.278	1.461	0.19
Poor quality roof	-0.184	0.146	-1.26	Miss Assets	0.822	0.956	0.86
Poor quality wall	-0.175	0.172	-1.02				
Animals	0.100	0.150	0.67	Constant	-0.464	0.650	-0.71
Land	0.407	0.142	0.29				
Number of Obs	1252			Pseudo R2	0.087		
LR Chi2 (56)	148.97			Log Likelihood	-776.97		
Prob>Chi2	0.000						

Figure 2.G.3.1: Estimated propensity scores (delay vs original treatment).



**Figure 2.G.3.2:** Stochastic dominance results (delay vs original treatment).**Table 2.G.3.6:** Propensity score matching: common support and number of observations in the common support (delay vs original treatment).

	Common support	Initial #	Delayed #	Band-width
IP	[0.341,0.785]	110	69	0.057
IL	[0.218,0.918]	226	152	0.066
NP	[0.191,0.859]	185	155	0.056
NL	[0.161,0.870]	206	145	0.060
Total		727	521	

Note: the acronyms refer to types : IP = Indigenous, Primary education; IL = Indigenous, Lower education; NP = Non-indigenous, Primary education; NL = Non-indigenous, Lower education.





### 3 | Distributional Effects of Oportunidades on Early Child Development

Published in *Social Science and Medicine*. Vol. 113, pp.42-49.

**Abstract:** The Mexican Oportunidades program is designed to increase human capital through investments in education, health, and nutrition for children in extreme poverty. Although the program is not expressly designed to promote a child's cognitive and non-cognitive development, the set of actions carried out by the program could eventually facilitate improvements in these domains. Previous studies on the Oportunidades program have found little impact on children's cognition but have found positive effects on their non-cognitive development. However, the majority of these studies use the average outcome to measure the impact of the program and thus overlook other "non-average" effects. This paper uses stochastic dominance methods to investigate results beyond the mean by comparing cumulative distributions for both children who are and children who are not aided by the program. Four indicators of cognitive development and one indicator of non-cognitive development are analyzed using a sample of 2,595 children aged two to six years. The sample was collected in rural communities in Mexico in 2003 as part of the program evaluation. Similar to previous studies, the program is found to positively influence children's non-cognitive abilities: children enrolled in the program manifest fewer behavioral problems compared with children who are not enrolled. In addition, different program effects are found for girls and boys and for indigenous and non-indigenous children. The ranges where the effect is measured cover a large part of the outcome's distribution, and these ranges include a large proportion of the children in the sample. With respect to cognitive development, only one indicator (short-term memory)

shows positive effects. Nevertheless, the results for this indicator demonstrate that children with low values of cognitive development benefit from the program, whereas children with high values do not. Overall, the program has positive effects on child development, especially for the most vulnerable, who are at the bottom of the distribution.

### 3.1 Introduction

Insufficient investments in human capital that are typically found among the poor seriously compromise children's present and future well-being. Children with poor health, nutrition, and education are less likely to develop the necessary skills for functioning in the economic and social realms in adulthood. Unhealthy conditions early in life, for example, have detrimental effects on the immunological system and development of the brain, and as a consequence, lead to poor cognition, problems with conduct, and difficulties in developing social relations at school (Walker, Wachs, Gardner, Lozoff, Wasserman, Pollitt, Carter, Group, et al. (2007), Walker, Chang, Powell, and Grantham-McGregor (2005), Chang, Walker, Grantham-McGregor, and Powell (2002)). Additionally, these conditions lead to low educational attainment and wage earnings (Duc, 2011; Case and Paxson, 2010; Schick and Steckel, 2010). The Mexican *Oportunidades* program aims to improve early life conditions by providing monetary transfers to disadvantaged families conditioned on regular investments in health, nutrition, and education. The program also provides scholarships to school-aged children (Fernald, Gertler, and Neufeld, 2008; Levy, 2006; Skoufias, 2005). Altogether, these investments are expected to break the intergenerational cycle of poverty.

Previous evaluations of *Oportunidades* and of similar interventions in other countries, generally show positive effects on the participants (Parker, Rubalcava, and Teruel, 2008; Ranganathan and Lagarde, 2012; Lagarde, Haines, and Palmer, 2009, 2007). For example, *Oportunidades* has been associated with greater height-for-age (Gertler, 2004; Behrman and Hoddinott, 2005; Farfán, Genoni, Rubalcava, Teruel, and Thomas, 2011), lower prevalence of anemia, and fewer episodes of illness (Gertler, 2004; Rivera, Sotres-Alvarez, Habicht, Shamah, and Villalpando, 2004). Additionally, positive effects on school enrollment and performance have been documented, both for primary and junior high school (Schultz, 2004; Behrman and Hoddinott, 2005). However, recent studies have drawn attention to the heterogeneity in these effects among participants: Todd and Winters (2011), who found that the program increases the probability of on-time school enrolment, report greater impacts for children with literate mothers, whereas Behrman and Hoddinott (2005) show better progression rates for girls throughout primary school. Heterogeneous effects of the program on health care utilization have also been detected between indigenous and non-indigenous participants (?), and evidence provided in ? points to greater impacts on the poorest in regard to the use of contraceptive methods.

Bearing in mind such heterogeneity, I investigate the effect of *Oportunidades* on children's cognitive and non-cognitive development using stochastic dominance methods which entail

the analysis of cumulative distributions. The procedure has the advantage of showing how the effects are distributed across children; thus, it addresses the possibility of disregarding important information for the analysis of the program. Most studies in the impact evaluation literature rely on some sort of average measure to assess the effect of an intervention. In fact, the analysis of the mean has become the standard rule in most impact evaluation studies. The mean effect, however, leads to skewed results when the program affects individuals in different ways. For example, a positive mean effect could be the result of few participants obtaining very large benefits, whereas many are hurt, as suggested by [Deaton \(2009\)](#). Looking at how effects are distributed could also address concerns about efficiency if the program brings fewer benefits to subgroups in greater need ([Gakidou, Oza, Fuertes, Li, Lee, Sousa, Hogan, Vander Hoorn, and Ezzati, 2007](#); [De Janvry and Sadoulet, 2006](#)). Few impact evaluation studies have employed distributions to assess the effect of interventions similar to *Oportunidades*, one study by [Djebbari and Smith \(2008\)](#) evaluates consumption patterns among *Oportunidades*' participants, finding evidence of systematic variation by subgroups. This suggests the varying effects could also exist in other domains where the program expects results. To my knowledge, this is the first paper that evaluates developmental outcomes for children in *Oportunidades* using non-average methods. Moreover, only few studies have used stochastic dominance criteria in the context of impact evaluation –Some examples are [Verme \(2010\)](#); [Naschold and Barrett \(2010\)](#), and more recently, ?. This latter study concentrates on health outcomes for children in *Oportunidades* and detects greater effects for the most deprived subgroups.

To illustrate the convenience of the analysis of distributions, I take as a point of departure the results of the three studies so far that have investigated the impact of *Oportunidades* on child development. The first two studies, by [Gertler \(2004\)](#) and by [Fernald, Gertler, and Neufeld \(2009\)](#), show a positive average effect on behavioral but not cognitive development after 3-6 and 10 years of exposure, respectively. The third study, by [Ozner, Fernald, Manley, and Gertler \(2009\)](#), concentrates on behavioral problems and shows how children exposed to the program for 3.5-5 years experience a reduction in aggressive/oppositional symptoms (but not anxiety/depressive symptoms). The results presented here suggest that some effects of the program on early child development have indeed been overlooked. In particular, cognitive effects not previously detected for boys and for children with non-indigenous background are now found. Overall, the results indicate that children with lower values of cognitive development benefit from the program, whereas those with higher values do not. Additionally, the majority of the children in the sample are consistently better off in terms of non-cognitive abilities.

## 3.2 Methods

I use stochastic dominance criteria to assess the effect of the program on five indicators of early child development. The procedure consists of comparing the cumulative distributions of each indicator for children in both a control group and a treatment group. A positive effect of the program occurs whenever the distribution of treated children dominates that of the control. Given the difference in composition between the treatment and control samples, I carry out a Propensity Score Matching procedure (PSM) to address the possible bias. The analysis is performed for the entire sample and for the sample split in two on the bases of gender and indigenous background of the children.

### 3.2.1 Data and Sample

The data were collected in 2003 by the *Oportunidades* staff in rural communities in Mexico as part of the external evaluation of the program. A subsample of children aged two to six years was selected for the purpose of assessing their cognitive and non-cognitive development. The sample contains information for two groups: children from families incorporated into the program in 2000 (treatment), and those from households not incorporated into the program (INSP, 2005).

The selection of the treatment sample proceeded as follows. In the first stage in 1997, highly deprived communities were identified and randomly assigned for program participation by *Oportunidades*' authorities. As a result, 186 localities with at least 500 inhabitants, and at most 2500, were selected during this stage. In the second stage, socioeconomic and demographic conditions were assessed to identify which households within the selected localities were eligible for the program. A marginality index based on income, demographic composition, and dwelling conditions of the household indicated whether a family was eligible for the program or not (Todd, 2004). Finally, a sample of children from eligible families was selected. Given that eligibility does not necessarily imply that these families agreed to participate, I only take into account those children for whom administrative records indicate their families received at some point monetary transfers from the program.

The selection of the control group, on the other hand, was conducted by a PSM procedure. This PSM is independent of the procedure followed in this article and it was conducted by the *Oportunidades*' authorities with the objective of selecting localities and not children, as in our case. Deprived localities where the program did not operate in 2003 were matched to treatment localities with similar observable characteristics (Todd, 2004). However, as explained by ?, this exercise has two problems. First, information on the set of characteristics

used to categorize deprived localities in 2003 comes from the National Census in 2000, when the treatment group had already entered the program. Second, localities instead of individuals were matched during this phase. Therefore, differences in terms of pre-program characteristics between treated and non-treated children might still exist.

To check if the comparability of the sample was compromised, I performed a logistic regression using as a dependent variable participation in the program (one indicating a child being enrolled in the program, zero otherwise) and as covariates a set of observable characteristics as shown in table 3.2 in appendix 3.A. As observed in the table, the head of the household in the treatment group was more likely to be older, male, and from an indigenous background, but less likely to be educated and to have a job in comparison with the control group. Additionally, differences in terms of the demographic composition, dwelling conditions, and the type and quality of the assets available in the household were detected. These results indicate that children in the control and treatment samples were not similar; thus, differences in child development might be due in part to environmental or socioeconomic conditions and not the result of *Oportunidades*. For instance, one could erroneously infer a positive effect if the treated children initially had higher developmental levels thanks to the higher education levels of their parents.

### 3.2.2 Propensity Score Matching

To overcome this problem, I performed a matching procedure at the individual level. The procedure consisted of weighting observations according to the probability that a child in the sample belongs to a household in the treatment group, i.e., according to the propensity of being enrolled in the *Oportunidades* program. The effectiveness of the PSM has been analyzed by Diaz and Handa (2006), who contrast the PSM with a perfect randomized experiment. They show that the PSM is reliable as long as a sufficiently large set of covariates is taken into account to generate the propensity score and the outcomes are measured in the same way for both the controls and treatments. The outcomes analyzed in this paper were collected using the same instrument for treatments and controls, and the set of covariates is very close to the specification used in other studies that analyze the *Oportunidades* program (Van de gaer, Vandenbossche, and Figueroa, 2014; Behrman, Parker, and Todd, 2009, 2011). All this suggests that the PSM can be used in this context to address the bias induced by unobserved factors.

The propensity score was calculated on the basis of the children's characteristics in 1997, when the program was not yet in place. For the treatment group, this information was obtained from baseline data collected in 1997 by the *Oportunidades*' authorities, and for the

control, the data were extracted from a set of retrospective questions on the 1997 households' characteristics collected in 2003 (INSP, 2005). The balance of the resulting weighted sample was once again tested by a logistic regression using the same set of covariates previously mentioned. The results, presented in table 3.4 in appendix 3.B, indicate that the resulting weighted sample is balanced in terms of observable characteristics.

After calculating the propensity score, the sample was split in two according to the children's genders and ethnicities. Previous studies have identified heterogeneous effects among these demographic groups. The division of the sample thus has the purpose of detecting possible differential effects across these groups. The final step consisted of matching observations between children in the treatment group and control group based on the propensity score. The matching was conducted for the complete sample and for each one of the demographic groups, i.e., boys in the control group were matched with boys in treatment, indigenous children in the control group were matched with indigenous children in treatment, and so forth. The stochastic dominance analysis was then performed for each of the five matched groups. Appendix 3.B provides a detailed explanation of the matching procedure. For additional information on the weighting procedure, see appendix 3 in ?. For more information on the PSM method, see Blundell and Dias (2009). The number of children in the original sample and after the PSM is provided in table 3.1. As can be observed, there are fewer children in each of the groups after the matching procedure because observations with very high or very low propensity scores could not be matched and, therefore, were not taken into account during the final analysis. The final sample for the analysis, then, is the matched sample, which is provided in the third and fourth columns in the table.

**Table 3.1:** Observations before and after PSM

	Original sample		Matched sample	
	Treatment	Control	Treatment	Control
Entire sample	1,087	1,625	1,086	1,509
Boys	541	813	541	752
Girls	546	812	539	756
Indigenous	450	374	405	361
Non-indigenous	635	1,241	634	1,144

### 3.2.3 Stochastic Dominance

I follow Van de gaer, Vandenbossche, and Figueroa (2014), who use the same test of stochastic dominance to assess the effect of an intervention. Imagine that  $\theta$  represents a random

variable that measures a child’s development such that a larger value of  $\theta$  represents a higher level of development. The level of development is measured through outcomes that can be of a cognitive or non-cognitive nature. Consider two groups of children that, before being incorporated into the program, are similar in terms of a set of observable socioeconomic, demographic, and environmental characteristics,  $X$ . Except for their status in the program (treatment or control), children are assumed to be comparable according to these characteristics. The characteristics composing  $X$  are assumed to be determinants of child development. Additionally, define the cumulative distribution function (CDF) of  $\theta$  for children in the program as  $F^T(\theta|X)$  and for children out of the program as  $F^C(\theta|X)$ . If  $F^C(\theta|X) \geq F^T(\theta|X)$  for all  $\theta$  and  $F^C(\theta|X) > F^T(\theta|X)$  for some  $\theta$ , then  $F^T(\theta|X)$  first-order stochastically dominates  $F^C(\theta|X)$ , and a positive effect of the program can be established. In other words, first-order dominance implies that at each percentile in the distribution, children in the program achieve higher developmental levels than children in the control.

To establish dominance, the following test, proposed by Davidson and Duclos (2013), was performed: Let  $U \subseteq \Theta$  be the union of the supports of the cumulative distributions of children not enrolled in the program (control) and children in the program (treatment), respectively  $F^C(\theta|X)$  and  $F^T(\theta|X)$ . We test the null hypothesis of nondominance of  $F^C(\theta|X)$  by  $F^T(\theta|X)$ ,

$$\max_{\theta \in U} \left( F^T(\theta|X) - F^C(\theta|X) \right) \geq 0,$$

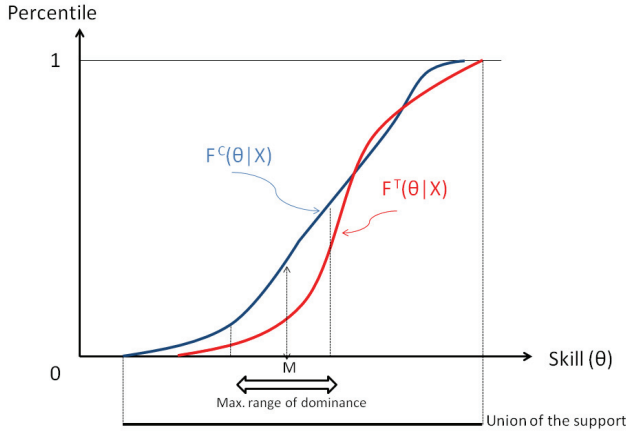
against the alternative hypothesis that  $F^T(\theta|X)$  first-order stochastically dominates  $F^C(\theta|X)$ ,

$$\max_{\theta \in U} \left( F^T(\theta|X) - F^C(\theta|X) \right) < 0.$$

The test identifies the maximum range over the distribution where dominance exists by first detecting the point where the difference between both curves is greatest and is statistically significant (Point “M” in figure 3.1). The test reports the largest range around this point where the null can be rejected at a certain level of significance. The procedure thus gives the maximum interval where dominance exists, that is, an interval where one group of children has higher levels of development in comparison to children in the other group. If the effect is in favor of the program, the group that has higher levels is the treatment group.

In summary, given the differences observed between the treatment and control groups in terms of pre-program characteristics, a propensity score matching was implemented. The PSM procedure consists of weighting observations based on the probability of being treated, which balances the sample in terms of observable characteristics. The balanced sample was then divided into two demographic groups according to gender and ethnicity, and the weighted



**Figure 3.1:** Test of first-order stochastic dominance

observations in the treatment group and the corresponding non-weighted observations in the control group were used to construct the cumulative distributions needed for the stochastic dominance test. The effect of the program is eventually obtained by comparing the distribution of cognitive and non-cognitive outcomes for treated and non-treated children using stochastic dominance criteria.

### 3.3 Outcomes

I analyze four indicators of cognitive development. The first indicator was constructed using the results of the Spanish version of the *Peabody Picture Vocabulary Test* (PPVT) –a test for measuring vocabulary recognition that has been used extensively as a proxy of language development among pre-schooled children aged 3-6 years (See the study of Paxson and Schady (2007) in Ecuadorian children and the study of Gertler and Fernald (2004) for children in *Oportunidades*). Children selected for the study were exposed to cards each containing four images. Next, the interviewer said a word that the children were asked to identify among the set of images in front of them. If the answer was correct, another set of images corresponding to a more difficult word was presented. This procedure continued until the child made six mistakes in eight consecutive questions or until the set of images was completed. Given that

no norms for children with similar characteristics and age ranges are available, I use raw scores to construct the cumulative distribution functions. Possible bias due to differences in age among the children was also taken into account by including age in months in the calculation of the propensity score.

The second measure of cognitive development is the Spanish version of the *MacArthur Communicative Development Inventories* (CDI). The test evaluates communication skills of children aged 24-35 months and is designed to assess early language milestones in young Spanish-speaking children. The test asks parents to identify words and phrases that their children know (understand and pronounce) from a list. The total number of words selected by the parents is then summed and used as an index of language ability. In comparison with similar tests, the CDI has been found to be more effective in assessing early language development (Jackson-Maldonado and Bárcenas Acosta, 2006). The fact that previous studies have documented higher scores in the CDI test for children receiving supplemental nutrition reinforces the hypothesis that the children in the program could also score higher in comparison with children in the control group (Gertler and Fernald, 2004; O'Connor, Hall, Adamkin, Auestad, Castillo, Connor, Connor, Fitzgerald, Groh-Wargo, Hartmann, et al., 2001).

Finally, non-cognitive development was assessed through the *Achenbach Child Behavior Checklist* (Achenbach Index). The test evaluates early behavioral and socio-emotional development among children aged 24-72 months. Based on parental responses, a child's problems, such as hyperactivity, bullying, bad conduct, violence at home, and responsiveness were directly rated by the parents according to a list where these problems were exposed to parental scrutiny. As in the CDI case, the index is constructed by summing the number of positive answers given by the parents, but in contrast with the other four indicators, larger values are indicative of poor development. The test has previously been used to assess the effects of parental background and environmental conditions on the behavioral development of children (Kahn, Brandt, and Whitaker, 2004; Pachter, Auinger, Palmer, and Weitzman, 2006).

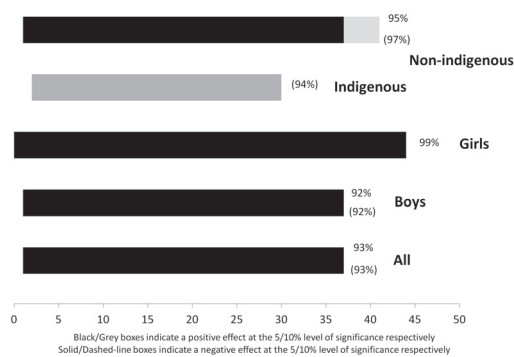
### 3.4 Results

The results of the tests of non-dominance for the entire sample and for each of the groups are presented in figures 3.2 through 3.6. Each figure contains boxes that indicate the range where an effect is found, i.e., where dominance exists. The observed range where each outcome is defined is shown at the bottom of each figure (horizontal axis). Black and grey boxes indicate a positive effect at the 5 and 10% level of significance, respectively. A positive effect occurs whenever the CDF of the treatment group dominates the corresponding CDF of the control

group (except for the Achenbach Index, in which case a higher value indicates more behavioral problems and therefore a worse outcome). A negative effect, on the other hand, is indicated by a white box (solid line) if the rejection occurs at the 5% level, and by a dashed white box if the rejection is at the 10%. Because it is easier to reject the null hypothesis at the 10% level of significance, the ranges are larger in comparison with ranges when the null is tested at the 5% level. The numbers appearing in parentheses next to each box show the proportion of the population included in the range where the effect occurs. If the number appears in parentheses, it refers to the effect at the 10% level of significance.

As observed in figure 3.2, the effect of the program on non-cognitive development is positive throughout the distribution. This is in line with previous studies finding that the program decreases children’s behavioral problems, but these results show that the ranges where the null of non-dominance can be rejected cover almost the entire variable’s distribution. At the 5% level of significance, the effect is favorable for more than 90% of the population in the entire sample and in each of the groups (indicated with a black or grey box).

**Figure 3.2:** Stochastic dominance results: Achen Index (Behavior problems)

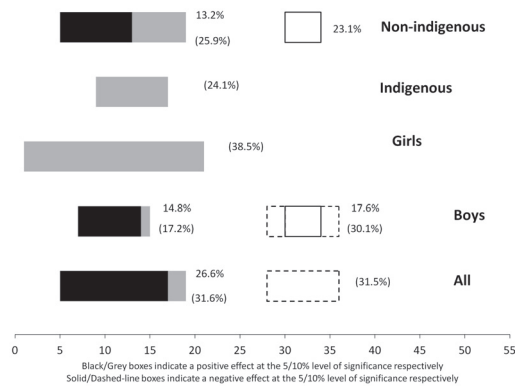


Figures 3.3 and 3.4 show the results for two cognitive indicators. In comparison with the evidence on non-cognitive development (figure 3.2), the results for short-term memory ability in Figure 3 are mixed. On the one hand, for the full sample, the effect is positive for children with lower values of cognitive ability (in black and grey). On the other hand, there is a negative effect for 17.6% of boys and 23.1% of non-indigenous children with higher values for cognitive ability (indicated by a white box). Additionally, there is a positive effect for approximately one-third of the girls in the sample at the 10% level of significance and for 15%

of the boys at the 5% level of significance. Additionally, the results on short-term memory emphasize the convenience of the methodology because it allows for the identification of the part of the distribution in which the effect is present. Here, it is possible to observe positive effects for boys and non-indigenous children, but only for those scoring low on the test.

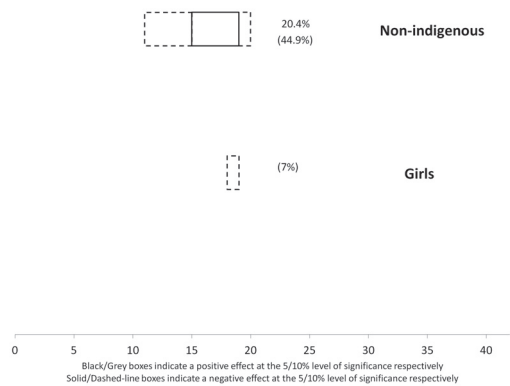
Results for visual integration are presented in Figure 3.4. *Oportunidades* has no positive effects for the complete sample or for any of the groups. On the contrary, the results show that the control group outperforms the treatment group for a small fraction of girls (7%) in a small range at the 10% level of significance. Similarly, approximately one-fifth of non-indigenous children in control group score better when the test is performed at the 5% level of significance, whereas the proportion increases to close to 50% when the test is rejected at the 10% level of significance.

Figure 3.3: Stochastic dominance results: Woodcock-Johnson (short-term memory)



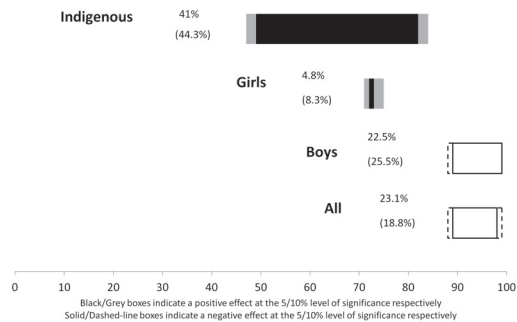
Finally, the results for the two measures of language development are shown in figures 3.5 and 3.6. The CDI is higher for 5% of girls and for 41% of boys enrolled in the program, and negative effects are observed for 22.5% of boys in the sample. In general, there are no positive effects of the program for the CDI index when looking at the full sample, whereas negative effects are observed for 23.1% of the population at the 5% level of significance. Similarly, no positive effects are observed when looking at the *Peabody Picture Vocabulary Test* in figure 3.6, and the effects are negative for the majority of the population (80.1%), but only at the 10% level of significance. The observation of such effects across groups shows that detrimental effects appear for girls and indigenous children at the 10% level of significance and that these effects are lessened because the ranges are smaller in comparison with the results for the

**Figure 3.4:** Stochastic dominance results: Woodcock-Johnson (Visual integration)



complete sample.

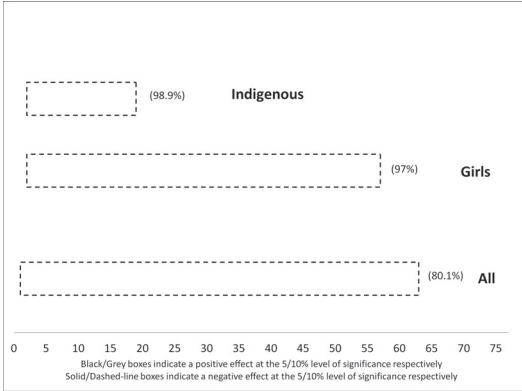
**Figure 3.5:** Stochastic dominance results: Communicative Development Inventories (CDI)



### 3.5 Sensitivity Analysis

The results presented above are in line with previous evidence on non-cognitive development; however, the results for short-term memory ability are surprising because none of the previous studies detected effects on this dimension. I argue in this paper that such discrepancies are

**Figure 3.6:** Stochastic dominance results: Peabody Picture Vocabulary Test (PPVT)



due to the inability of more traditional methods to detect effects along the distribution. Nonetheless, such differences could be driven by differences in the composition between the samples used in these studies and the sample analyzed here. To address this issue, below I provide the results of three additional analyses using samples that are comparable to those used in the aforementioned studies. Additionally, I present average and quantile treatment effects (QTE) to set the benchmark against which the stochastic dominance results are to be compared. The average effect is extensively reported in impact evaluation studies and is also the approach followed in previous studies that analyze the effect of the program on child development. In addition, QTE has been proposed as an alternative to the average approach to gain insight on effects occurring beyond the mean and, as such, provides a good comparison with the stochastic dominance approach. The estimator implemented is the unconditional QTE proposed by [Firpo \(2007\)](#).

The first of these analyses relies on a sample that is very similar to the one used in [Gertler and Fernald \(2004\)](#). The results, presented in the first part of appendix [3.C](#), are comparable to those obtained in the main analysis: children in the program show better results for non-cognitive ability than children in control group, and the percentages where dominance occurs are similar to those previously observed. Similarly, the results for short-term memory ability are positive for children with low cognition and negative for children with higher values. The similarity between these results and those obtained in the main analysis presented above reinforces the assumption that differences in the results obtained with the analysis of distributions are not driven by the composition of the sample, but by the

methodology. On the other hand, the second analysis focuses on the same sample of children used by Ozner, Fernald, Manley, and Gertler (2009), and, in line their analysis, the results are in favor of the program. However, the sample used in this second analysis was not balanced in terms of observable characteristics (after implementing a PSM), therefore, the possibility of biased results exists.

In addition, a third analysis addresses the possibility that the results are affected by the inclusion of children aged two in the sample. In contrast with children aged 3-6, these children had been exposed to the program for less time to the program by the time the data were collected in 2003. Therefore, there is a possibility that the observed results are driven, at least in part, by the inclusion of this group. The third analysis in the appendix shows the results of the stochastic dominance test using only children aged two in the sample. The analysis was conducted using the Achenbach Index because information on the other four outcomes was not collected for two-year-old children. As can be observed, the results do not substantially differ from the main findings of this paper: children in the program are consistently better off in terms of non-cognitive skills than children in the control group. The ranges where dominance occurs are also very similar for all groups, as is the proportion of children included in these ranges. This finding suggests that results obtained using stochastic dominance criteria are not affected by the time of exposure for the children in the sample.

Finally, average treatment effects and QTE are presented in tables 3.5 and 3.6, respectively, at the bottom part of appendix 3.C. The propensity scores used to generate these results were the same as those previously used for the stochastic dominance analysis; thus, the same set of covariates were included in all cases. The analysis of the mean shows that although all children seem to benefit from the program in terms of non-cognitive development, in general, no effects on cognition were found. The results only show impacts on the cognitive skills of indigenous children, namely, on short-term memory skills. However, these results may undermine the heterogeneity in the effect among children that might occur in this case because the benefits are not homogenous across families (the amount of the monetary transfer varies depending on the number of children attending school, for example). The stochastic dominance approach addresses this problem; however, most evaluation studies addressing the same issue rely on QTE, another popular methodology that goes beyond the mean effect. QTE takes into account these concerns by comparing cumulative distributions; however, this approach differs from the stochastic dominance criterion used in this article in two respects. First, QTE provide the average difference at a particular percentile, and the test of stochastic dominance indicates ranges where the CDFs are significantly different. Second, QTE requires selecting *a priori* the percentiles at which the effect is to be measured. Although QTE shows

a broader perspective of the effect of the program along the distribution, some effects at percentiles not selected could still be ignored, which is not the case with the test of stochastic dominance. The results for the quantile treatment analysis in table 3.6 show positive effects on non-cognitive ability, but contrary to the stochastic dominance results, the effect is only statistically significant for girls and non-indigenous children. Additionally, children in the lowest quantiles of short-term memory ability have positive values, whereas children in higher quantiles have negative ones. These results are in line with the stochastic dominance analysis, but these effects are statistically significant only in a few cases. Note, however, that although the stochastic dominance approach and QTE are intended for looking at effects along the distribution, both could lead to different results because the latter requires pre-selecting the quantiles at which one tests whether the program has an effect, whereas the former gives intervals where the difference between both CDFs is statistically significant.

### 3.6 Discussion

The results indicate that *Oportunidades* improves early child development in both cognitive and non-cognitive domains. Virtually all children in the program whose families have been incorporated into the program for approximately three years show higher values in the index of non-cognitive development. Likewise, the effect on short-term memory is positive for children with low cognitive skills, whereas the effect is negative for those with higher levels. Such differential effects are not observable with more traditional methods that rely on the mean effect or effects at specific quantiles, as in the case of QTE. In contrast, stochastic dominance criteria permit the assessment of the effect of the program along the entire distribution. The combination of stochastic dominance and PSM methods strengthen the results presented here, given the pre-program differences in the composition of the treatment and control groups that are found in the sample.

Although the program does not explicitly address skill development, the package of services—consisting of monetary transfers; in-kind assistance on health, nutrition, and education; and information provided by the program (e.g., educational workshops on health and nutrition for mothers)—could improve a child’s readiness to develop these skills. The results on these domains are well established (Parker, Rubalcava, and Teruel, 2008); however, how the program affects children’s skill development is less clear. The evidence offered in previous studies suggests scant effects on cognitive development and some positive results on non-cognitive development. The results presented here, in contrast, suggest that the scope of these studies is limited because important effects beyond the mean and for different groups are overlooked.



For example, Ozner, Fernald, Manley, and Gertler (2009) find no differences between genders and ethnicities, and Gertler and Fernald (2004) find effects only for girls and not for boys.

More specifically, results on non-cognitive ability are found in ranges that cover a substantial part of the distribution and for the majority of children in the sample. Additionally, substantial effects for both boys and girls are observed. These results are not trivial if one acknowledges the importance of the so-called “soft skills” and the role they have for functioning in daily adult life. As noted by Heckman and Kautz (2012), personality traits (non-cognitive abilities) are highly predictive of success in life, and contrary to cognitive skills, they remain throughout life, even if no further investments are made. Therefore, interventions that foster the development of non-cognitive skills of children should occupy a central role in the design of anti-poverty interventions.

Regarding the effect of the program on cognition, the results are quiet surprising because none of the studies mentioned before find effects of the program in any of the outcomes analyzed in this paper. Unlike these studies, the results offered here suggest that *Oportunidades* affects one dimension of cognitive development, namely, short-term memory ability. As mentioned by Gertler and Fernald (2004), a child’s short-term memory ability depends more on factors such as better nutrition and health –which the program most directly supports. In contrast, other indicators that show no effects and depend more on parental stimuli are in areas where the program is less involved. Nonetheless, these authors find only a weak mean effect on short-term memory skill, which illustrates the advantage of the method proposed here. For example, there are positive effects for one-third of the girls in the sample, at the 10% level of significance. Additionally, the program is favorable for 15% of the boys, but disadvantageous for another 17% (figure 3.3). Such opposite effects could have been “mixed” in Gertler and Fernald (2004), and expressed merely as a weak mean effect. Additionally, the results suggest that the program is effective in helping the most vulnerable children because those in the lower part of the distribution of short-term memory ability are better off.

Indeed, although effects measured through average and quantile treatment analysis are consistent with those measured through stochastic dominance criteria, the latter method leads to more significant results and deepens the analysis within and across groups. A drawback of the methodology, however, is the impossibility of translating the results in terms of magnitudes. The test of stochastic dominance only shows the part of the distribution where the effect exists and the proportion of the children that are affected by the program. Perhaps a combination of stochastic dominance criteria with methods more suitable for providing magnitudes would be more adequate. An additional limitation is the lack of clarity about the negative results on the higher part of the distribution when looking at short-term memory

skills. The nature of the data makes clarification on this point difficult, but traditional models of parental altruism –in which parents invest less in their children when returns are low– can explain why cognitive skills can decrease as a result of the program (Becker, 1981).

Similarly, more research is needed to unveil the mechanism through which *Oportunidades* affects children’s development. Some possible paths have been previously suggested: families using the monetary transfers to buy more and healthier food, medicines, and goods that could positively affect children’s well-being (Fernald, Gertler, and Neufeld, 2008). The program could also have improved children’s development by providing services such as medical and nutritional attention and information on positive habits. Additionally, parents could have experienced less stress thanks to these improvements and eventually pay more attention to and provide better care for their children (Gertler and Fernald, 2004). However, the design of the evaluation of the program does not allow a disentangling the specific effect of each one of these domains. A solution to this problem would be the estimation of a structural model. However, this exercise is beyond the scope of the present analysis.

## Bibliography

- BECKER, G. S. (1981): *A Treatise on the Family*. Harvard University Press, Cambridge, MA.
- BEHRMAN, J., AND J. HODDINOTT (2005): “Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition,” *Oxford Bulletin of Economics and Statistics*, 67(4), 547–569.
- BEHRMAN, J., S. PARKER, AND P. TODD (2009): “Medium-term impact of the oportunidades conditional cash transfer program on rural youth in Mexico,” in *Poverty, Inequality and Poverty in Latin America*, ed. by S. Klasen, and N. F, pp. 219–270. MIT Press, Cambridge.
- BEHRMAN, J., S. PARKER, AND P. E. TODD (2011): “Do conditional cash transfers generate lasting benefits? a five year follow-up of PROGRESA/Oportunidades,” *Journal of Human Resources*, 46(93-122).
- BLUNDELL, R., AND M. C. DIAS (2009): “Alternative approaches to evaluation in empirical microeconomics,” *Journal of Human Resources*, 44(3), 565–640.
- CALIENDO, M., AND S. KOPEINIG (2008): “Some practical guidance for the implementation of propensity score matching,” *Journal of economic surveys*, 22(1), 31–72.
- CASE, A., AND C. PAXSON (2010): “Causes and consequences of early-life health,” *Demography*, 47(1), S65–S85.
- CHANG, S., S. WALKER, S. GRANTHAM-MCGREGOR, AND C. POWELL (2002): “Early childhood stunting and later behaviour and school achievement,” *Journal of Child Psychology and Psychiatry*, 43(6), 775–783.
- DAVIDSON, R., AND J. DUCLOS (2013): “Testing for Restricted Stochastic Dominance,” *Econometric Reviews*, 32(1), 84–125.
- DE JANVRY, A., AND E. SADOULET (2006): “Making conditional cash transfer programs more efficient: designing for maximum effect of the conditionality,” *The World Bank Economic Review*, 20(1), 1–29.
- DEATON, A. S. (2009): “Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development,” Discussion paper, National Bureau of Economic Research.

- DIAZ, J., AND S. HANDA (2006): “An Assessment of Propensity Score Matching as a Non-experimental Impact Estimator: Evidence from Mexico’s PROGRESA Program,” *Journal of Human Resources*, 41(2), 319–345.
- DJEBBARI, H., AND J. SMITH (2008): “Heterogeneous impacts in PROGRESA,” *Journal of Econometrics*, 145(1), 64–80.
- DUC, L. T. (2011): “Height and cognitive achievement of Vietnamese children,” *World Development*, 39(12), 2211–2220.
- FARFÁN, G., M. E. GENONI, L. RUBALCAVA, G. TERUEL, AND D. THOMAS (2011): “Oportunidades and its impact on child nutrition,” Discussion paper, Working Paper.
- FERNALD, L. C., P. J. GERTLER, AND L. M. NEUFELD (2008): “Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico’s Oportunidades,” *The Lancet*, 371(9615), 828–837.
- (2009): “10-year effect of Oportunidades, Mexico’s conditional cash transfer programme, on child growth, cognition, language, and behaviour: a longitudinal follow-up study,” *The Lancet*, 374(9706), 1997–2005.
- FIRPO, S. (2007): “Efficient semiparametric estimation of quantile treatment effects,” *Econometrica*, pp. 259–276.
- GAKIDOU, E., S. OZA, C. V. FUERTES, A. Y. LI, D. K. LEE, A. SOUSA, M. C. HOGAN, S. VANDER HOORN, AND M. EZZATI (2007): “Improving child survival through environmental and nutritional interventions: the importance of targeting interventions toward the poor,” *Journal of the American Medical Association*, 298(16), 1876–1887.
- GERTLER, P. (2004): “Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA’s Control Randomized Experiment,” *American Economic Review*, 94(2), 336–341.
- GERTLER, P. J., AND L. C. FERNALD (2004): “The medium term impact of Oportunidades on child development in rural areas,” *National Institute of Public Health*.
- HECKMAN, J. J., AND T. KAUTZ (2012): “Hard evidence on soft skills,” *Labour economics*, 19(4), 451–464.
- INSP (2005): “General Rural Methodology Note,” *Instituto Nacional de Salud Pública*, INSP2005.

- JACKSON-MALDONADO, D., AND R. BÁRCENAS ACOSTA (2006): "Assessing Early Communicative Abilities in Spanish-Speaking Children from Low-SES Families.," *Zero to Three (J)*, 27(1), 22–27.
- KAHN, R. S., D. BRANDT, AND R. C. WHITAKER (2004): "Combined effect of mothers' and fathers' mental health symptoms on children's behavioral and emotional well-being," *Archives of Pediatrics & Adolescent Medicine*, 158(8), 721–729.
- LAGARDE, M., A. HAINES, AND N. PALMER (2007): "Conditional cash transfers for improving uptake of health interventions in low-and middle-income countries: a systematic review," *Jama*, 298(16), 1900–1910.
- (2009): "The impact of conditional cash transfers on health outcomes and use of health services in low and middle income countries," *Cochrane Database Syst Rev*, 4(4).
- LEVY, S. (2006): *Progress against poverty: Sustaining Mexico's Progres-Oportunidades program*. Brookings Institution Press, Washington, DC.
- NASCHOLD, F., AND C. BARRETT (2010): "A Stochastic Dominance Approach to Program Evaluation with an Application to Child Nutritional Status in Kenya," Working Paper.
- O'CONNOR, D. L., R. HALL, D. ADAMKIN, N. AUESTAD, M. CASTILLO, W. E. CONNOR, S. L. CONNOR, K. FITZGERALD, S. GROH-WARGO, E. E. HARTMANN, ET AL. (2001): "Growth and development in preterm infants fed long-chain polyunsaturated fatty acids: a prospective, randomized controlled trial," *Pediatrics*, 108(2), 359–371.
- OZNER, E. J., L. C. FERNALD, J. G. MANLEY, AND P. J. GERTLER (2009): "Effects of a conditional cash transfer program on children's behavior problems," *Pediatrics*, 123(4), e630–e637.
- PACHTER, L. M., P. AUINGER, R. PALMER, AND M. WEITZMAN (2006): "Do parenting and the home environment, maternal depression, neighborhood, and chronic poverty affect child behavioral problems differently in different racial-ethnic groups?," *Pediatrics*, 117(4), 1329–1338.
- PARKER, S., L. RUBALCAVA, AND G. TERUEL (2008): "Evaluating Conditional Schooling and Health Programs," in *Handbook of Development Economics*, ed. by T. Schultz, and J. Strauss, vol. 4, pp. 3963–4035. Elsevier.

- PAXSON, C., AND N. SCHADY (2007): “Cognitive development among young children in Ecuador the roles of wealth, health, and parenting,” *Journal of Human resources*, 42(1), 49–84.
- RANGANATHAN, M., AND M. LAGARDE (2012): “Promoting healthy behaviours and improving health outcomes in low and middle income countries: a review of the impact of conditional cash transfer programmes,” *Preventive medicine*, 55, S95–S105.
- RIVERA, J., D. SOTRES-ALVAREZ, J. HABICHT, T. SHAMAH, AND S. VILLALPANDO (2004): “Impact of the Mexican Program for Education, Health, and Nutrition (PROGRESA) on Rates of Growth and Anemia in Infants and Young Children,” *The Journal of the American Medical Association*, 291(21), 2563–2570.
- SCHICK, A., AND R. STECKEL (2010): “Height as a Proxy for Cognitive and Non-Cognitive Ability,” *NBER Working Paper N 16570*.
- SCHULTZ, P. (2004): “School subsidies for the poor: evaluating the Mexican Progresa poverty program,” *Journal of Development Economics*, 74(1), 199–250.
- SILVERMAN, B. W. (1986): *Density Estimation for Statistics ad Data Analysis*. Chapman & Hall, London.
- SKOUFIAS, E. (2005): *PROGRESA and its impacts on the welfare of rural households in Mexico*, vol. 139. Intl Food Policy Res Inst.
- TODD, J. E., AND P. WINTERS (2011): “The effect of early interventions in health and nutrition on on-time school enrollment: Evidence from the Oportunidades program in rural Mexico,” *Economic Development and Cultural Change*, 59, 549–581.
- TODD, P. (2004): “Design of the evaluation and method used to select comparison group localities for the six year follow-up evaluation of Oportunidades in rural areas,” Discussion paper, International Food Policy Research Institute.
- VAN DE GAER, D., J. VANDENBOSSCHE, AND J. L. FIGUEROA (2014): “Children’s health opportunities and project evaluation: Mexico’s Oportunidades program,” *The World Bank Economic Review*, 28, 282–310.
- VERME, P. (2010): “Stochastic Dominance, Poverty and the Treatment Effect Curve,” *Economics Bulletin*, 30(1), 365–373.

- WALKER, S. P., S. M. CHANG, C. A. POWELL, AND S. M. GRANTHAM-MCGREGOR (2005): “Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: prospective cohort study,” *The Lancet*, 366(9499), 1804–1807.
- WALKER, S. P., T. D. WACHS, J. M. GARDNER, B. LOZOFF, G. A. WASSERMAN, E. POLLITT, J. A. CARTER, I. C. D. S. GROUP, ET AL. (2007): “Child development: risk factors for adverse outcomes in developing countries,” *The lancet*, 369(9556), 145–157.





# Appendices

## 3.A Characteristics of the households in the sample in 1997

**Table 3.2:** Characteristics of the households in the sample in 1997. Logistic regression with dependent variable as 1 if observation belongs to treatment, 0 otherwise.

Variable	Coefficient	SE*	z	Variable	Coefficient	SE*	z
Height mother	−0.021	0.009	−2.24	Draft animals	0.17	0.114	1.49
PPVT score mother	−0.001	0.002	−0.46	Land ownership	0.531	0.106	5.01
age of child (months)	0.004	0.003	1.09	Blender	0.176	0.133	1.33
Age HH head	0.013	0.007	1.91	Fridge	−0.112	0.194	−0.58
Age spouse	0.011	0.007	1.53	Gas stove	−0.358	0.146	−2.44
Sex HH head	2.215	0.352	6.3	Gas heater	−0.655	0.363	−1.81
Indigenous HH head	0.664	0.274	2.42	Radio	0.592	0.1	5.87
Indigenous spouse	−0.256	0.278	−0.92	Hifi	0.351	0.252	1.39
Educ HH head	0.244	0.115	2.13	TV	0.643	0.119	5.41
Educ spouse	0.41	0.118	3.46	Video	−0.515	0.345	−1.49
Work HH head	−1.085	0.264	−4.11	Wash machine	0.038	0.329	0.11
Work spouse	−0.611	0.163	−3.76	Car	−1.213	0.468	−2.59
# Children 0-5 years	0.087	0.048	1.81	Truck	−0.208	0.285	−0.73
# Children 6-12 years	0.206	0.042	4.91	Guerrero	0.537	0.19	2.83
# Children 16-20 years	0.025	0.074	0.33	Hidalgo	0.945	0.211	4.49
# Children 13-15 years	0.166	0.084	1.96	Michoacán	0.646	0.179	3.6
# Women 20-39 years	0.025	0.12	0.21	Puebla	1.085	0.15	7.21
# Women 40-59 years	−0.022	0.156	−0.14	Querétaro	−0.131	0.22	−0.6
# Women 60+ years	−0.02	0.184	−0.11	San Luis Potosí	0.458	0.154	2.97
# Men 20-39 years	−6.777	0.236	−2.87	Missing Age Spouse	4.332	0.716	6.05
# Men 40-59 years	−0.367	0.161	−2.27	Missing Indg Hh head	−0.421	1.776	−0.24
# Men 60+ years	−0.677	0.236	−2.87	Missing Indg Spouse	1.878	1.695	1.11
# Rooms	0.007	0.047	0.15	Missing Work Hh head	−3.094	1.669	−1.85
Electricity	−0.017	0.115	−0.15	Missing Work Spouse	−3.593	1.637	−2.19
Running water land	−0.885	0.115	−7.67	Missing water land	−1.155	1.735	−0.67
Running water house	0.435	0.209	2.08	Missing water house	−0.634	0.824	−0.77
Dirtfloor	−0.118	0.119	−0.99	Missing height mother	−3.135	1.444	−2.17
Poor quality roof	0.004	0.109	0.04	Missing PPVT score mother	0.225	0.299	0.75
Poor quality walls	0.471	0.127	3.71	Constant	−0.787	1.47	−0.53
Number of observations	2,712			Pseudo R2	0.1951		
LR Chi2(57)	712.61			Log Likelihood	−1469.78		
Prob. > chi2	0						

\*Standard errors adjusted at locality level

### 3.B Matching procedure

The matching procedure was carried out in several steps. In a first step, the propensity scores for each observation were calculated through a logistic regression using the set of covariates presented in table 3.2. The balancing property of the propensity score was tested using Stata 10; according to this test, the optimal number of blocks was 9. In total, 522 tests were performed (one test for each one of the 58 covariates in each block) and only in 8 cases the null hypothesis of the balancing test was rejected.

In a second step, the propensity score matching procedure was implemented within each of the four groups selected for the analysis and for the entire sample, i.e., the matching was performed between boys in treatment with boys in control, indigenous in treatment with indigenous in control, and so forth. The scores are plotted and presented in figure 3.8 below. Note that the ranges of the propensity scores for each group are provided on the bottom of each figure. Also, a trimming procedure based on the “Max Min” criterion proposed by [Caliendo and Kopeinig \(2008\)](#) was implemented. The resulting trimmed samples were then used to carry out a Kernel matching based on an Epanechnikov kernel function. Table 3.3 shows the resulting common supports as well as the number of observations before and after the trimming procedure. The first two columns contain the original number of observations while the third and fourth columns show the resulting observations after the trimming and matching procedures. The ranges of the common support ranges are provided in the fifth column. Additionally, the optimal bandwidth parameter used in the kernel function is shown in the last column of the table. The parameter was calculated by the formula  $\alpha = 1.06 \min(\sigma, \frac{\rho}{1.34})$  presented in [Silverman \(page 45-47 1986\)](#).

Finally, in order to assess the balance of the covariates used to generate the propensity score, an additional logistic regression using the matched sample is presented in table 3.4. As can be observed in the table, only 5 out of 58 covariates are statistically significant. These results suggest that both groups are very similar in terms of these set of pre-program characteristics.

**Table 3.3:** Propensity scores and number of observations by group

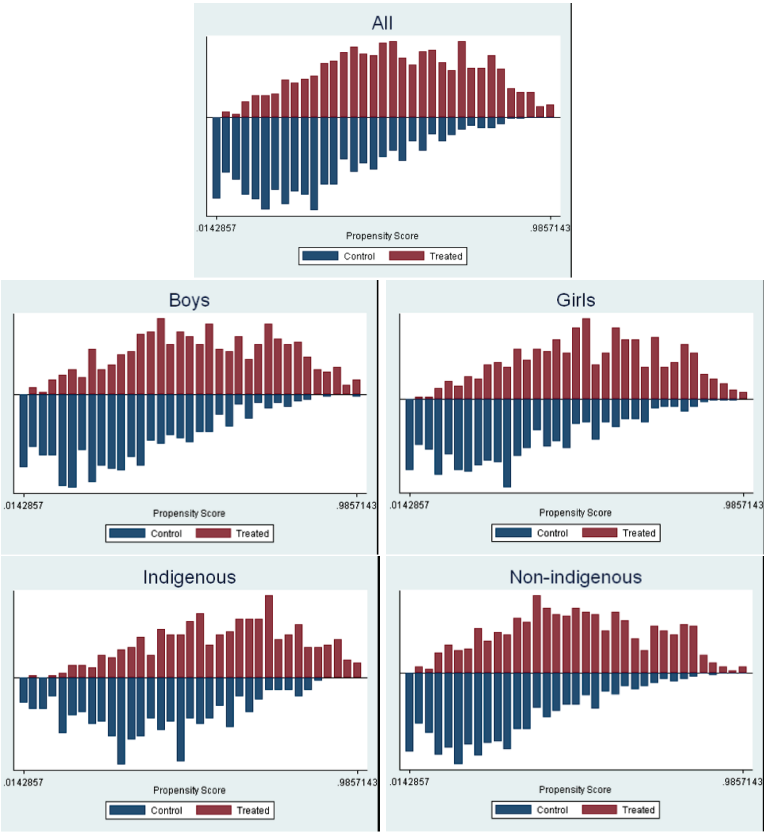
	# observations		# observations in common support		Common Support	Optimal bandwidth ( $\alpha$ parameter)
	Treatment	Control	Treatment	Control		
All	1,087	1,625	1,086	1,509	[0.0448, 0.9847]	0.242
Boys	541	813	541	752	[0.0472, 0.9847]	0.242
Girls	546	812	539	756	[0.0448, 0.9445]	0.239
Indig.	450	374	405	361	[0.0477, 0.8787]	0.22
Non- ind.	635	1,241	634	1,144	[0.0448, 0.9847]	0.23

**Table 3.4:** Characteristics of the households in the matched sample. Logistic regression with dependent variable as 1 if observation belongs to treatment, 0 otherwise.

Variable	Coefficient	SE*	Z	Variable	Coefficient	SE*	Z
Height mother	-0.002	0.01	-0.16	Draft animals	0.117	0.118	0.99
PPVT score mother	0	0.003	0.17	Land ownership	0.041	0.112	0.37
Age of child (months)	0.003	0.004	0.94	Fan	0.035	0.2	0.17
Age HH head	0.009	0.007	1.22	Blender	0.134	0.139	0.96
Age spouse	-0.007	0.007	-0.91	Fridge	0.056	0.21	0.27
Sex HH head	0.265	0.363	0.73	Gas stove	-0.244	0.155	-1.57
Indigenous HH head	0.181	0.293	0.62	Gas heater	-0.322	0.4	-0.81
Indigenous spouse	-0.029	0.302	-0.1	Radio	0.087	0.107	0.82
Educ HH head	0.02	0.119	0.17	Hifi	0.116	0.263	0.44
Educ spouse	0.099	0.126	0.78	TV	0.223	0.123	1.82
Work HH head	-0.632	0.241	-2.62	Video	-0.077	0.356	-0.22
Work spouse	-0.185	0.179	-1.03	Wash machine	0.009	0.342	0.03
# Children 0-5 years	0.043	0.05	0.85	Car	0.08	0.434	0.18
# Children 6-12 years	0.019	0.044	0.43	Truck	-0.243	0.308	-0.79
# Children 16-20 years	0.015	0.079	0.19	Guerrero	0.297	0.197	1.51
# Children 13-15 years	0.161	0.09	1.79	Hidalgo	0.571	0.206	2.77
# Women 20-39 years	0.034	0.125	0.27	Michoacán	0.131	0.194	0.67
# Women 40-59 years	0.023	0.165	0.14	Puebla	0.154	0.158	0.98
# Women 60+ years	0.11	0.187	0.59	Querétaro	-0.016	0.234	-0.07
# Men 20-39 years	0.051	0.116	0.44	San Luis Potosí	0.249	0.166	1.5
# Men 40-59 years	-0.123	0.176	-0.7	Missing Age Spouse	-0.681	0.891	-0.76
# Men 60+ years	-0.012	0.242	-0.05	Missing Indg Hh head	0.453	1.756	0.26
# Rooms	-0.012	0.048	-0.26	Missing Indg Spouse	1.017	1.714	0.59
Electricity	-0.064	0.124	-0.52	Missing Work Hh head	-0.672	1.651	-0.41
Running water land	-0.118	0.123	-0.96	Missing Work Spouse	-0.068	1.6	-0.04
Running water house	0.049	0.229	0.21	Missing water land	0.854	1.319	0.65
Dirtfloor	-0.019	0.126	-0.15	Missing water house	0.106	0.847	0.13
Poor quality roof	-0.012	0.115	-0.1	Missing height mother	-0.292	1.565	-0.19
Poor quality walls	0.219	0.132	1.66	Missing PPVT score mother	0.034	0.314	0.11
				Constant	-0.275	1.582	-0.17
Number of observations	2,595			Pseudo R2	0.021		
Wald chi2(58)	67.56			Log Likelihood	-1760.94		

\*Standard errors adjusted at locality level

**Figure 3.7:** Estimated Propensity Scores by group



3.C Sensitivity analysis

Analysis #1

Analysis using a sample similar to the one used in [Gertler and Fernald \(2004\)](#). Like in their analysis, two groups were used to construct the treatment sample: a group enrolled in 1998 (treatment-1998) and a second one incorporated two years later (treatment-2000). Both groups were compared with the control group in 2003. The following tables show the results of the test of stochastic dominance for the 5 outcomes described in the article.

Figure 3.8: Stochastic dominance results: Achen Index (Behavior problems)

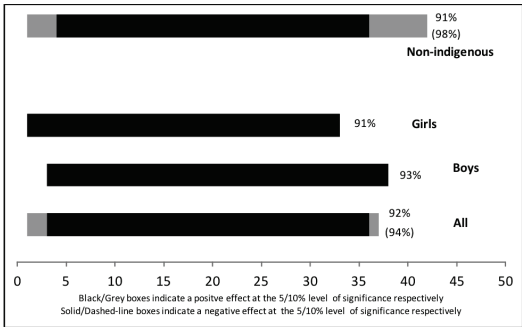


Figure 3.9: Stochastic dominance results: Woodcock-Johnson (short-term memory)

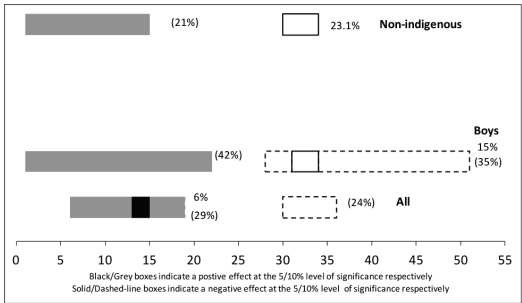


Figure 3.10: Stochastic dominance results: Woodcock-Johnson (Visual integration)

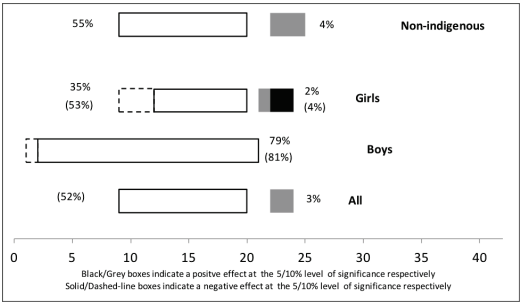


Figure 3.11: Stochastic dominance results: Communicative Development Inventories (CDI)

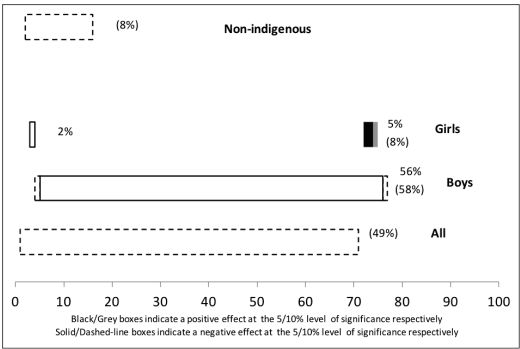
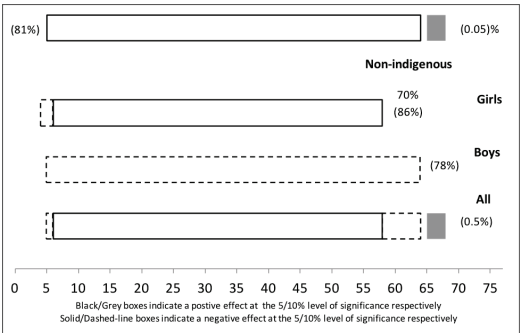


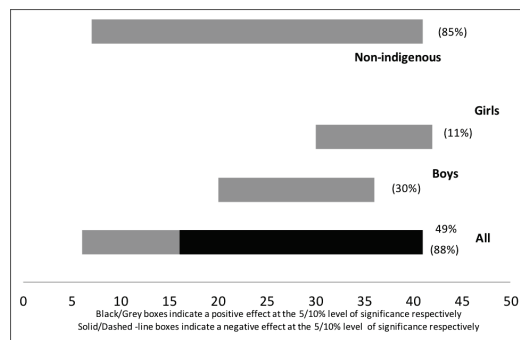
Figure 3.12: Stochastic dominance results: Peabody Picture Vocabulary Test (PPVT)



## Analysis # 2

Analysis using a similar sample to the one used by Ozner, Fernald, Manley, and Gertler (2009). The results are presented in the following figure:

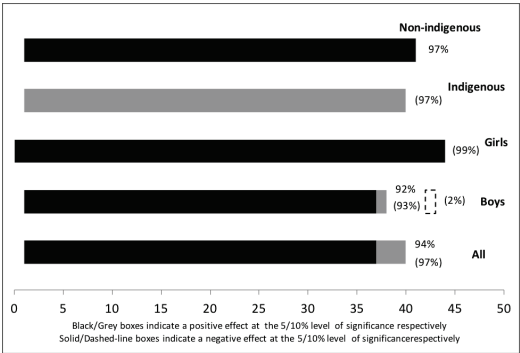
**Figure 3.13:** Stochastic dominance results: Achen Index (Behavior problems)



## Analysis # 3

Finally the following figure shows the results for children aged 2 in the sample. The original sample contains children aged 2-6 years in 2003. The families of treated children were incorporated in 2000. Therefore, children between 3 and 6 years old in 2003 were all exposed to the program 3 years. This is not the case for children aged 2 in 2003 that were exposed less time. If time of exposure is important, then children in this age-group might respond differently to the program in comparison with older children. The analysis was carried out using the Achen Index (non-cognitive indicator) because this is the only outcome that contains information for this group, the rest contain information for children older than 3 years old. Note that the CDI was also collected for children aged 2, however, this analysis is already presented together with the main results. As can be observed from the graph, the results do not substantially change: the effect is positive for all groups and in a large part of the distribution. Moreover, the ranges and the percentages that indicate the proportion of treated children are very similar. This evidence suggests that the main results including children aged 2 are not affected by the time of exposure of these children.

Figure 3.14: Stochastic dominance results for children aged 2 years (Achen Index)





Average and Quantile Treatment Effects

Table 3.5: Average mean effect and descriptive statistics

	Mean	SD	Range	Average effect <sup>(1)</sup>	SE <sup>(2)</sup>	t
<b>ALL</b>						
Achen Index	18.327	10.108	[0,50]	-1.435	0.378	-3.796
Short-term memory	21.372	10.995	[0,54]	0.087	0.536	0.163
Visual integration	10.184	6.72	[0,42]	-0.289	0.289	-1
CDI	65.672	28.33	[0,100]	1.65	2.576	0.623
Peabody	14.313	12.113	[0,77]	-1.153	0.629	-1.834
<b>Boys</b>						
Achen Index	18.804	10.273	[0,50]	-0.948	0.711	-1.334
Short-term memory	21.182	11.136	[0,54]	-0.259	0.632	-0.411
Visual integration	10.209	6.822	[0,31]	-0.428	0.369	-1.16
CDI	64.628	28.986	[0,100]	2.853	3.871	0.737
Peabody	14.743	12.01	[0,77]	-0.788	0.663	-1.188
<b>Girls</b>						
Achen Index	17.847	9.921	[0,46]	-1.963	0.609	-3.222
Short-term memory	21.562	10.851	[0,51]	0.435	0.648	0.671
Visual integration	10.16	6.617	[0,42]	-0.157	0.366	-0.428
CDI	66.814	27.599	[1,100]	0.129	3.923	0.033
Peabody	13.89	12.204	[0,73]	-1.42	0.875	-1.623
<b>Indigenous</b>						
Achen Index	18.687	10.159	[0,46]	-1.108	0.672	-1.647
Short-term memory	17.431	10.328	[0,43]	2.078	0.799	2.601
Visual integration	8.637	6.457	[0,33]	0.615	0.492	1.25
CDI	63.5	27.685	[0,100]	3.327	4.894	0.68
Peabody	10.799	8.771	[0,57]	-0.064	0.793	-0.081
<b>Non-Indigenous</b>						
Achen Index	18.105	10.032	[0,50]	-1.784	0.492	-3.624
Short-term memory	23	10.859	[0,54]	0.273	0.601	0.455
Visual integration	10.819	6.721	[0,42]	-0.345	0.322	-1.073
CDI	66.453	28.616	[0,100]	0.729	3.96	0.184
Peabody	15.794	12.924	[0,77]	-0.594	0.76	-0.782

<sup>(1)</sup> Estimation using Kernel Matching method; average treatment on the treated reported

<sup>(2)</sup> Bootstrapped standard errors

Table 3.6: Quantile treatment effects at selected deciles<sup>(1)</sup>

	<i>q</i> = 0.1	<i>q</i> = 0.2	<i>q</i> = 0.3	<i>q</i> = 0.4	<i>q</i> = 0.5	<i>q</i> = 0.6	<i>q</i> = 0.7	<i>q</i> = 0.8	<i>q</i> = 0.9
<b>All</b>									
Achen Index	-1	-1	-1	-1	-2*	-1	-1	-1	-1
Short-term memory	1	1	1	0	-2	-1	0	-1	-1
Visual integration	0	0	-1	0	0	-1	-1	-1	0
CDI	2	4	10	7	6	5	3	0	1
Peabody	0	-1*	0	-2*	-1	-1	-1	-1	-2
<b>Boys</b>									
Achen Index	-1	-1	0	0	0	0	-1	0	-1
Short-term memory	0	2	0	-1	-2	-1	0	-1	-1
Visual integration	0	-1	-1	-1	-1	-2	-1	0	1
CDI	0	9	11	5	6	6	4	1	2
Peabody	0	-1	-1	-1	-1	-1	-1	0	0
<b>Girls</b>									
Achen Index	-2**	-2**	-2**	-3***	-4***	-3***	-2	-2	-1
Short-term memory	2*	2	2	0	-1	-2	-1	0	-1
Visual integration	0	1	0	0	0	0	-1	-1	-2
CDI	4	4	6	10	8	4	3	0	-2*
Peabody	-1***	-1	-1	-1	-2	-2	-2	-3	-4
<b>Indigenous</b>									
Achen Index	-1	0	1	-1	-1	-1	-1	0	0
Short-term memory	4	2	2	2	2	1	1	0	1
Visual integration	0	0	0	0	0	0	1	0	1
CDI	4	1	7	12	12	8	3	-6	-4
Peabody	0	-1	-1	0	-1	-2	-2	-2	1
<b>Non-indigenous</b>									
Achen Index	-1	-2*	-1	-2**	-3***	-2**	-2	-2	-1
Short-term memory	2	1	1	-1	-1	-1	-1	-1	-1*
Visual integration	0	0	0	0	0	-2*	-1	-1	0
CDI	-1	1	8	3	5	5	4	0	1
Peabody	0	0	-2*	0	0	0	-1	0	-2

<sup>(1)</sup> Bootstrapped standard errors  
\*, \*\* and \*\*\* indicate statistical significance at the 10%, 5% and 1%, respectively.

## 4 | Did Progresá Reduce Inequality of Opportunity for School Re-enrollment?

with Dirk Van de gaer

**Abstract:** The way school enrollment is distributed over children facing different circumstances (like gender, indigenous origin, parental education or place of residence) should be an important concern in the evaluation of social programs like Progresá. The literature has shown that Progresá increases average school enrollment for children aged 12 and above, especially during the transition from primary to secondary school, and some authors found evidence that the program has larger effects for some groups, for instance for girls than for boys. This paper finds inspiration in the recent literature on the measurement of inequality of opportunity to evaluate the effect of the program, taking its effect on inequality of opportunity in a systematic way into account. Focussing on school re-enrollment opportunities for each grade attained, our findings are that Progresá improved aggregate opportunities, average school re-enrollment and reduced inequality of opportunity for school re-enrollment. Moreover, using the Human Opportunity Index, proposed by de Barros, Ferreira, Molinas Vega, and Chanduvi (2009) de Barros, Vega, and Saavedra (2008), and used extensively in Molinas, de Barros, Saavedra, and Guigale (2012), or, the Gini Opportunity Aggregator, the decrease in inequality of opportunity accounts for about 15 to 40 % of the total effect of the program on aggregate opportunities.

## 4.1 Introduction

A key concern in developing countries is the provision of access to basic opportunities for all (see, e.g., [Molinas, de Barros, Saavedra, and Guigale \(2012\)](#)). One such opportunity is access to education. When universal access is not achieved, the next best thing is that all children have equal access at the highest possible level. As powerfully advocated by both economists and political philosophers (see, e.g. [Roemer \(1993\)](#) [Roemer \(1998\)](#) [Fleurbaey \(2008\)](#) or [Cohen \(1989\)](#)), a child's access to education should not be influenced by his circumstances, defined as characteristics over which he has no control, which include his family background, race, gender and place of residence. Hence not only the level of children's opportunities, but also the way these opportunities are distributed between children with different circumstances is highly relevant.

Many social interventions, including Progresá, are motivated by the concern to provide ample opportunities for all. The question whether such social interventions influence both the level of children's opportunities and their distribution is key, and has been analyzed in Chapters 1 and 2. Mexico's Progresá, in 2002 extended in reach and renamed Oportunidades, and in 2014 changed into Prospera<sup>1</sup>, aims at expanding children's educational opportunities. In fact, the program is explicitly designed to give parents incentives to keep their children in school by increasing school subsidies when the children are in a higher grade (to make schooling more attractive compared to child labor) and by providing larger grants for girls than for boys at ages when otherwise girls are more likely to drop out.

Previous evaluations find positive effects of the program: they identify, for instance, the effect of the program on average school enrollment of children of a particular age. There is widespread agreement in the evaluation literature that Progresá increased average school enrollment, with large effects for children aged 12 and above (see, e.g., [Schultz \(2004b\)](#); [Behrman, Sengupta, and Todd \(2005\)](#); [Todd and Wolpin \(2006\)](#); [Attanasio, Meghir, and Santiago \(2012\)](#); [Dubois, de Janvry, and Sadoulet \(2012\)](#)). Important conclusions follow from these studies. [Schultz \(2004b\)](#), on page 22, for instance, reports that: "the impact ... is to increase the educational attainment of a cohort of poor youth by 0.66 years of schooling ..., for which the youth earn a 12% higher wage per year of schooling over their adult working lifetimes (age 18-65) based on the 1996 urban wage structure."

Some studies go further and identify average effects for children having different circumstances. [Schultz \(2004b\)](#) and [Todd and Wolpin \(2006\)](#), for instance, find larger positive treatment effects on school enrollment for girls than for boys. [Figueroa \(2014\)](#) goes beyond

---

<sup>1</sup>See also the description of the program in the general introduction of the dissertation

average treatment effects for different groups and identifies the effect on the entire distribution of child cognitive and non-cognitive skills for girls, for boys, for children from indigenous and non-indigenous origin.<sup>2</sup> Van de gaer, Vandenbossche, and Figueroa (2014) explicitly focus on children's circumstances, measured by the four possible combinations based on whether they had at least one indigenous parent or not and whether they had at least one parent that completed primary education or not.<sup>3</sup> They establish whether children's health improves or not, as a function of their circumstances and their place in the distribution of the health outcome. While such studies identify which types of children gain from the program and which ones do not, their non-parametric methodology limits the number of different circumstances they can consider and they do not provide a systematic assessment of the consequences of the program on inequality of opportunity.

This paper provides such an assessment. We draw on recent contributions in the literature on inequality of opportunity that propose to measure the inequality in counterfactual distributions which only reflect differences in circumstances (see for example, Pistolesi (2009) and Ferreira and Gignoux (2011))<sup>4</sup>. As in the previous two chapters, program participation is taken as a circumstance, because it lies beyond children's control. In our case, we construct two counterfactuals: one in case children are treated and one in case they are not. To this end, we first estimate a logistic regression to obtain the effect of the program on school re-enrollment for children facing different circumstances. Next, we use this regression to predict for each child, given his circumstances, the probability of being re-enrolled in school if living in a household that participates in the program, and the probability in case of living in a non-participant household. This gives us a vector of predicted re-enrollment probabilities, with and without the program, that can be used to compare the inequality contained in each of them.

To determine the effect of the program on inequality of opportunity, we provide, in first place, results for Generalized Lorenz and Lorenz dominance. Under certain conditions discussed below, one can unambiguously interpret Generalized Lorenz dominance as an improvement for all opportunity aggregators. Similarly, under certain conditions discussed, one can unambiguously interpret Lorenz dominance as an improvement for all inequality of opportunity measures. We also want to see how important the effect on inequality of opportunity is for the total effect on the aggregator. To this end, we use the Human Opportunity Index (de Barros, Ferreira, Molinas Vega, and Chanduvi, 2009; de Barros, Vega, and Saavedra,

---

<sup>2</sup>See Chapter 2.

<sup>3</sup>See Chapter 1.

<sup>4</sup>For an overview of this rapidly expanding literature, see Ferreira H.G. and Peragine (2015), Ramos and Van de gaer (2015) or Roemer and Trannoy (2014).

2008) and, as an alternative, the Gini Opportunity Aggregator. Both can be decomposed into changes that are due to changes in average re-enrollment (measuring how many opportunities for education are provided) and changes in an index that measures the inequality of the distribution of re-enrollment between children with different circumstances, which therefore measures changes in inequality of opportunity.

We find that, generally, Progresa leads to a distribution of school re-enrollment probabilities (conditional on circumstances) that both Generalized and Lorenz dominates the distribution of school re-enrollment in the absence of the program. For children that completed primary school (grade six) the results are particularly strong. Using the Human Opportunity Index or the Gini Opportunity Aggregator, the decrease in inequality of opportunity is responsible for about 15 to 40 % of the increase in aggregated opportunities.

## 4.2 Description of Progresa and sample selection

Progresa (Programa de Educación, Salud y Alimentación) was implemented by the Mexican government in 1997 for poor rural households. Its goal was to alleviate poverty and break the intergenerational transmission of poverty through the development of human capital. The program consists of two components. First, households receive cash transfers. Educational grants are conditional on children's school enrollment and on regular school attendance (an attendance rate of 85 per cent is required) and grants for consumption of food are conditional on regular medical check-ups in health clinics and attendance of health and nutrition talks. Second, in-kind health benefits and nutritional supplements are provided for children up to age five, and for pregnant and lactating women. The analysis presented here refers to the first year after the program began<sup>5</sup>. The reason, as explained in more detail below, is that during the first two years of the program, experimental data was available to evaluate the effects of participation. After 2 years, the group that was randomly selected as a control, was incorporated in the program, and the experimental design of the program was lost.

Primary education in Mexico consists out of 6 grades, followed by 3 years of secondary education and 3 years of upper secondary or high school. The educational grants started at the third grade in primary school, increased through grade levels and ended after secondary school<sup>6</sup>. At the secondary level, girls received higher grants than boys – Skoufias (2005) Table 1.1 for the exact amounts. The total monthly monetary payments a household can receive (in the form of educational grants and grants for food) is capped. Between November 1998 and

<sup>5</sup>The analysis refers thus to the “immediate” and “delayed” groups described in the General Introduction of the dissertation and used in the present analysis.

<sup>6</sup>After 2001, grants were provided until the third grade of high school.

October 1999, households with school-aged children received per month on average 101 pesos through food related transfers and 139 through school related transfers. In total, between November 1998 and October 1999, these transfers represented more than 18 per cent of the total average value of household consumption in rural areas – Skoufias (2005), Table 1.5.

Based on data from national censuses, highly deprived rural localities with access to a primary school and a health clinic were identified. In these localities households that experienced extreme poverty (measured on the basis of household income, characteristics of the head of household, and variables related to dwelling conditions) were included in the program. For logistical reasons, not all localities eligible for participating in the program were enrolled at the same time. A random procedure assigned some localities to receive immediate treatment in November 1997 and monetary payments from May 1998; the others began receiving treatment in December 2000<sup>7</sup>.

Collecting data to make a rigorous evaluation of Progresa possible was an important concern of the program designers from the outset, and the delayed incorporation of some localities was instrumental to obtain these data. The data on children's education we use were collected in October 1998, when a first group of localities, hereafter referred as to immediate treatment group, already received treatment, but not a second groups of localities which was incorporated later (referred to as delayed treatment group). Data were collected in 320 immediate treatment localities and 186 delayed treatment localities, resulting in our treatment and control sample, respectively. Hence, we follow much of the literature (see, e.g., Gertler (2004); Schultz (2004a); Behrman, Sengupta, and Todd (2005); Todd and Wolpin (2006); Attanasio, Meghir, and Santiago (2012); Dubois, de Janvry, and Sadoulet (2012)) and use the delayed treatment sample as a control group in order to identify the short-run effects of the program<sup>8</sup>. In both treatment and control samples, we only consider the households that were eligible for the program; thus only the poor households of the selected localities are considered. In addition, we use baseline data collected in 1997 to obtain some of the information on children's background characteristics such as parental education, gender of the head of the household, as well as dwelling characteristics and demographic composition of the household. The reason for using the 1997 survey is the lack of background information in the 1998 sample. A caveat of such a procedure, however, is the difficulty to merge information from both years for some of the children observed in 1998. Our sample, thus, contains children whose background information in 1997 could be successfully traced with their information on

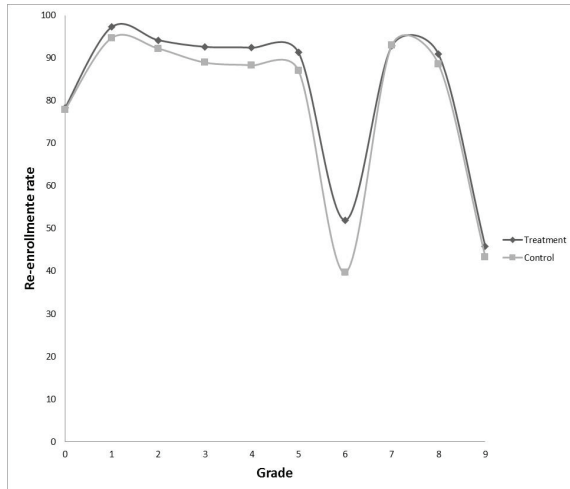
<sup>7</sup>See the Introduction of the dissertation for more details.

<sup>8</sup>Children that belong to treated households have been receiving treatment only for the last 8 months. Moreover, the sample is restricted to school-aged children, none of which can have benefited directly from the nutritional supplements (only given to children up to age 5). Hence the long run effects of the program are probably larger.

education from 1998<sup>9</sup>. For these reasons, average results presented in Section 4.4.1 refer to intend-to-treatment analysis.

Figure 1 below, provides for both immediate and delayed treatment samples, the percentage of the children that attained grade 0 to 9 as their highest completed grade and were re-enrolled two months after the academic year 1998-99 started in August 1998. For grade 0, this means that children did not succeed after attending the first year and were enrolled again in first grade in October 1998.

**Figure 4.1:** Re-enrollment rates in October 1998 per grade for immediate and delayed treatment sample.



*Source: Author's calculation.*

School re-enrollment is high across all grades, except for grade 6, which is when primary education is finished, and grade 9, which is when secondary school is finished. There are clear differences between the immediate and delayed treatment samples for children that completed grades 3 to 6. Especially for the latter, re-enrollment in the treatment sample is higher than in the control sample (delayed treatment).

Two remarks must be made at this point. First, a simple comparison of immediate and delayed treatment samples might be misleading if these samples differ in terms of observable or unobservable characteristics. Suppose, for instance that the immediate treatment sample con-

<sup>9</sup>Information on education was collected only for children between 6 and 17 years old in 1998. In total, 84% of these children have also information in 1997.



tains more children from higher educated parents and that higher educated parents keep their children longer in school. If this is the case, some of the difference in average re-enrollment between the initial and delayed treatment samples would be due to this higher frequency of educated parents in the immediate sample and not only be a consequence of the treatment, and the causal effect of the program would be unidentified. The consensus is, however, that the differences between these samples in terms of characteristics are limited (see, [Behrman and Todd \(1999\)](#)).<sup>10</sup> Second, looking at average re-enrollment per grade is not informative about the effect the program has on the distribution of re-enrollment between children confronted with different circumstances. For example, an increase in the average re-enrollment rate could result from some children having advantaged circumstances (for instance those with highly educated parents) seeing their re-enrollment rate increase much more than average, and those with disadvantaged circumstances (those having poorly educated parents), seeing their re-enrollment rate decrease. Since the ultimate goal of this paper is to evaluate the effect of Progresá on inequality of re-enrollment opportunities, we have to look at distributional effects. The next section describes in detail the procedure we follow to account for the effect of the program on inequality of opportunity.

## 4.3 Methodology

### 4.3.1 Theoretical framework

We have a fixed set  $N = \{1, \dots, n\}$  of children. For each child  $i$  we have determined  $p_i$ , the probability that he or she re-enrolls given his or her circumstances, where circumstances are characteristics of the child (or its environment) over which it has no control, such that we do not want to hold the child responsible for these characteristics. The higher a child's  $p_i$ , the more advantage the child's circumstances offer<sup>11</sup>. Therefore, in the literature on the measurement of inequality of opportunity,  $p_i$  is considered to measure the value of the child's opportunities. The vector  $p = (p_1, \dots, p_n) \in [0, 1]^n$  records the value of the opportunities of the children in our set  $N$ . To evaluate the opportunities of the children, we define an opportunity aggregator function  $O : [0, 1]^n \rightarrow \mathbb{R}$ .

We impose the following standard properties on this aggregator function. First,  $O(p)$  satisfies anonymity: any permutation of the vector  $p$  yields the same value for the opportunity

<sup>10</sup>In addition, we test for each grade, the balance of our sample using baseline data collected in 1997. The results, shown in the Appendix 4.A (Tables A.3–A.5) suggest that there are no substantial differences between the immediate and the delayed samples.

<sup>11</sup>This is a measurement statement, not a causal statement. Suppose that children facing some circumstances are more motivated to re-enroll at school. In that case, the effect of circumstances on their motivation and thereby on their re-enrollment probability will be attributed to their circumstances.

aggregator function as the original  $p$ . Second,  $O(p)$  satisfies non-decreasingness in every  $p_i$ : whenever there is one child whose opportunities increase, without decreasing another's opportunities, aggregate opportunities do not decrease. Third,  $O(p)$  satisfies the weak transfer principle: increasing the value of opportunities for a child with low opportunities by a given amount and decreasing the value of opportunities for a child with higher opportunities by the same amount, without giving the former more opportunities than the latter, cannot decrease the value of  $O(p)$ . Let  $p^N$  be the vector of opportunities in the absence of the program, and  $p^T$  the vector of opportunities with the program. We now have the following well-known result (Shorrocks, 1983).

**Proposition 1:** for all opportunity aggregator functions  $O(p)$  that satisfy anonymity, non-decreasingness and the weak transfer principle, aggregate opportunities in the vector  $p^T$  will not be smaller than in the vector  $p^N$  if and only if the Generalized Lorenz curve of  $p^T$  lies nowhere below the Generalized Lorenz curve for  $p^N$ .

Generalized Lorenz dominance is a condition under which we can unambiguously compare, for all opportunity aggregator functions satisfying the three properties, what happens to the aggregated opportunities of the children. We are, however, also interested in what happens to the inequality in the distribution of children's opportunities. Define an inequality measure as a function  $I : [0, 1]^n \rightarrow \mathbb{R}$ . We impose standard properties on this inequality measure. First,  $I(p)$  satisfies anonymity: any permutation of the vector  $p$  yields the same value for the inequality index as the original  $p$ . Second,  $I(p)$  satisfies the weak transfer principle: increasing the value of opportunities for a child with low opportunities by a given amount and decreasing the value of opportunities for a child with higher opportunities by the same amount, without giving the former more opportunities than the latter, cannot increase inequality. Third,  $I(p)$  is relative: multiplying all children's opportunities by the same positive constant does not affect the value of the inequality measure. We can now use the following standard result (Foster and Shorrocks, 1988).

**Proposition 2:** for all inequality of opportunity measures  $I(p)$  that satisfy anonymity, the weak transfer principle and relativity, inequality of opportunity in the vector  $p^T$  will not be larger than inequality of opportunity in the vector  $p^N$  if the Lorenz curve of  $p^T$  lies nowhere below the Lorenz curve of  $p^N$ .

We are also interested to decompose the effect of Progresá on aggregate opportunities into an effect on average opportunities and an effect on inequality of opportunity. This kind of question can be analyzed using "abbreviated" opportunity aggregator functions (by analogy

with “abbreviated social welfare functions” - see (Lambert, 2001)). More in particular, by defining average opportunities as  $C = (1/n) \sum_{i=1}^n p_i$ , we can write abbreviated opportunity aggregator functions as

$$O(p) = C [1 - I(p)]. \quad (4.1)$$

Following the procedure suggested by de Barros, Vega, and Saavedra (2008), the change in the values of abbreviated opportunity aggregator functions brought about by Progresá can be decomposed as follows:

$$\begin{aligned} \underbrace{O(p^T) - O(p^N)}_{\Delta_O} &= C^T(1 - I(p^T)) - C^N(1 - I(p^N)) \\ &= C^T(1 - I(p^T)) - C^N(1 - I(p^T)) + C^N(1 - I(p^T)) - C^N(1 - I(p^N)) \\ &= \underbrace{(C^T - C^N)(1 - I(p^T))}_{\Delta_C} + \underbrace{C^N(I(p^N) - I(p^T))}_{\Delta_I}. \end{aligned} \quad (4.2)$$

Equation (4.2) shows how the change in the Opportunity aggregator,  $\Delta_O$ , can be decomposed in two components:  $\Delta_C$ , the contribution of the change in average school re-enrollment, and  $\Delta_I$ , the contribution of the change in inequality of opportunity.

The Human Opportunity index, proposed by de Barros, Vega, and Saavedra (2008), takes as inequality measure the dissimilarity index

$$D = \frac{1}{2C} \frac{1}{n} \sum_{i=1}^n |p_i - C|, \quad (4.3)$$

which equals the percentage of average school re-enrollment that has to be taken from those with a school re-enrollment above average and given to those with a school re-enrollment below average, such that everyone, independent of his circumstances, ends up with the grade's average school re-enrollment (all  $p_i = C$ ). With the vector  $\hat{p}$  as the vector obtained after permuting the elements of  $p$  such that  $\hat{p}_1 \leq \hat{p}_2 \leq \dots \leq \hat{p}_n$  and  $m$  such that  $\hat{p}_m \leq C < \hat{p}_{m+1}$ , it is easy to show that the opportunity aggregator function corresponding to (4.1) and (4.3), the value of the Human Opportunity Index, can be written as

$$H = \frac{1}{n} \left[ \left(2 - \frac{m}{n}\right) \sum_{i=1}^m \hat{p}_i + \left(1 - \frac{m}{n}\right) \sum_{i=m+1}^n \hat{p}_i \right].$$

This aggregator function is strictly increasing in all  $\hat{p}_i$ . It also satisfies the weak transfer principle, but observe that transfers between children that have lower than average opportunities or transfers between children that have higher than average opportunities do not affect the

value of  $H$ . It only values transfers from children with above average opportunities to children with below average opportunities. This might be considered a disadvantage for those who defend the strong transfer principle, which requires that increasing the value of opportunities for a child with low opportunities by a given amount and decreasing the value of opportunities for a child with higher opportunities by the same amount, without giving the former more opportunities than the latter, always decreases inequality. There exist many ways of constructing an abbreviated opportunity aggregator function satisfying the strong transfer principle. We propose one, based on the familiar Gini index of inequality,

$$G = 1 + \frac{1}{n} - 2 \frac{\hat{p}_n + 2\hat{p}_{n-1} + 3\hat{p}_{n-2} + \dots + n\hat{p}_1}{n^2 C}. \quad (4.4)$$

It is easy to show that the opportunity aggregator function corresponding to (4.1) and (4.4), the value of the Gini Opportunity Aggregator, can be written as

$$S = \frac{1}{n^2} \sum_{i=1}^n (2i-1) \hat{p}_{n+1-i},$$

showing the familiar pattern of weights that are a linearly decreasing function of the rank order of the  $\hat{p}_i$ .

### 4.3.2 Selection of circumstances

In principle, to determine the extent of inequality of opportunity, we need a complete description of each child's circumstances. Our data contain a rich set of circumstances thanks to the information that was collected as part of the evaluation of the program in October 1998 and September 1997. Descriptive statistics of these variables are shown in Table A.1 in Appendix 4.A. The set of circumstances we used include personal and background characteristics and are gender, race (indigenous background), parental background (level of education of the parents), gender of the head of the household, whether his or her spouse lives in the same home and whether there was a secondary school in the locality where the child lives. These circumstances have been selected for the following reasons.

It is well established that large disparities still exist in Mexico in terms of access to and performance at school and part of such disparities are correlated with gender, girls being more disadvantaged than boys. To capture the effect of gender, we use a dummy indicating whether the child is a boy or a girl. Similarly, having an indigenous background is traditionally believed to diminish the possibilities that a child receives education given the social relegation of this part of the population. Indigenous background is measured by a variable that indicates if

the child belongs to a family where either the head of the household or the spouse of the head speaks an indigenous language. Speaking an indigenous language is perhaps the best indicator for indigenous background since it is not common that non-indigenous people in Mexico speak an indigenous language.

The effect of parental characteristics on children's schooling has been widely documented. Particularly important is the education of the parents. Molinas, de Barros, Saavedra, and Guigale (2012) show for example that parents' education is the most important circumstance to explain school enrollment of children aged 10 to 14 in most Latin American countries, including Mexico. In our sample, there are few household heads with more than secondary education (grades 7-9); therefore, educational background of the children is measured by dummy variables that indicate if the head of the household has no education, some education (without having completed primary school), or completed at least primary school. The same variables were constructed for the spouse of the head. Gender of the head of the household and whether the head and the spouse live in the same household are also indicated by dummies. The later helps to identify single parent families. The presence of a secondary school in the locality where the child lives is potentially important, as it could affect the decision of the child to go to primary or secondary school. Finally, as we have many missing values for some of the circumstances (see Part (b) of Table A.1) and to be able to use these observations in the estimation of the effects of the circumstances for which their values are not missing, additional dummies that indicate the presence of missing values are included as circumstances.

### 4.3.3 Empirical methodology

We need to estimate first the probability of re-enrollment for the children in the sample. For each grade obtained before the start of the school year<sup>12</sup>, we perform a logistic regression with school re-enrollment as a binary dependent variable ( $Y_i = 1$  if child  $i$  is enrolled in a particular grade and  $Y_i = 0$  otherwise), the  $K$  dimensional vector  $X_i$  as circumstances ( $X_{ik}$  is child  $i$ 's value for circumstance  $k$ ) and a dummy treatment variable  $T_i$  indicating whether the child participated in the program ( $T_i = 1$ ) or not ( $T_i = 0$ ). We also include interaction terms between circumstances and the treatment dummy variable. These interaction effects allow the treatment to have different effects for children with different circumstances (e.g.,

---

<sup>12</sup>Interviewees were asked for the last grade obtained in school in October 1998, when the academic year 1998-1999 had already began. Only information on children attending school was collected, therefore, we observe school re-enrollment.

girls versus boys or indigenous versus nonindigenous children). The specification is standard:

$$Prob(Y_i = 1) = \frac{\exp(\beta_0 + \sum_{k=1}^K \beta_k X_{ik} + \gamma_0 T_i + \sum_{k=1}^K \gamma_k T_i X_{ik})}{1 + \exp(\beta_0 + \sum_{k=1}^K \beta_k X_{ik} + \gamma_0 T_i + \sum_{k=1}^K \gamma_k T_i X_{ik})}. \quad (4.5)$$

The estimated values of the coefficients  $\beta_0, \beta_1, \dots, \beta_k$ , and  $\gamma_0, \gamma_1, \dots, \gamma_k$  are used to generate, for every child, two estimated probabilities:  $p_i^T$ , the predicted probability of re-enrollment if the child is treated and  $p_i^N$ , the corresponding probability if the child is not treated. Observe that, if two children, say  $i$  and  $j$ , have a different predicted probability ( $p_i^Z \neq p_j^Z$ ,  $Z = T$  or  $N$ ), this must be due to differences in circumstances. Hence, the inequality in the vector of predicted probabilities  $p_i^T$  (or  $p_i^N$ ) is entirely due to children having different circumstances, and is, for that reason, a measure of inequality of opportunity amongst children participating (not participating) in the program<sup>13</sup>. For a comparison with other measures, see Ramos and Van de gaer (2015). Average re-enrollment can be computed easily: with  $Z=T$  or  $N$ ,

$$C^Z = \frac{1}{n} \sum_{i=1}^n p_i^Z. \quad (4.6)$$

Having obtained the predicted values  $p_i^N$  and  $p_i^T$ , we rearrange them increasingly, into the vectors  $p^N$  and  $p^T$ , respectively. These vectors form the input in our evaluation exercise.

To construct the Generalized Lorenz curves, we define  $M_d^T$  as the value of the coordinate of the Generalized Lorenz curve in case of being treated and  $M_d^N$  as that in case of not being treated. The value of these coordinate at each decile  $d$  for  $d = 1, \dots, 10$ <sup>14</sup> is given by:

$$M_d^Z = \sum_{i=1}^{\tilde{d}(d)} p_i^Z, \quad (4.7)$$

where  $Z=T$  if treated or  $N$  otherwise and  $\tilde{d}(d)$  is the integer that is closest to  $[n/10] * d$ . For each  $d = 1, \dots, 10$ , the difference between the coordinates of the Generalized Lorenz curve in case of being treated and not being treated is  $M_d^D = M_d^T - M_d^N$ . Similarly, define, for each decile  $d = 1, \dots, 10$ ,  $L_d^T$  as the value of the coordinate of the Lorenz curve in case of being treated,  $L_d^N$  as that in case of not being treated and the difference between both as

<sup>13</sup>This is the basic idea of the ex-ante (direct) approach to the measurement of inequality of opportunity –see, e.g., Pistoiesi (2009) and Ferreira and Gignoux (2011). See also discussion in the General Introduction of the dissertation

<sup>14</sup>Given the computational burden that arises with the Bootstrap procedure explained in Appendix 4.B, we only look at values at each decile

$L_d^D = L_d^T - L_d^N$ . The value of the coordinates of the Lorenz curve at decile  $d$  is given by:

$$L_d^Z = \frac{1}{\tilde{d}(d)C^Z} \sum_{i=1}^{\tilde{d}(d)} p_i^Z, \quad (4.8)$$

where  $Z=T$  if treated or  $N$  otherwise,  $C^Z$  is the respective value of the average predicted probability  $p_i^Z$  as shown in (4.3.3) and  $\tilde{d}(d)$  is as above.  $M_d^D > 0$  for all  $d$  implies that the Generalized Lorenz curve of treated children dominates that of non-treated children, and therefore, that Progresca increases aggregate opportunities for all aggregator functions satisfying anonimity, the transfer principle and non-decreasingness (see Proposition 1). Similarly,  $L_d^D > 0$  for all  $d$  indicates Lorenz dominance such that Progresca reduces inequality of opportunity for all inequality indices satisfying anonimity, relativity and the transfer principle (see Proposition 2).

We can easily compute  $D^T$  and  $H^T$  by replacing  $p$  by  $p^T$  in (4.3) and (4.3.1), respectively. Similarly replacing  $p$  by  $p^N$  results in  $D^N$  and  $H^N$ . Following the decomposition procedure described in (4.2), the difference in the Human Opportunity Index that is due to Progresca can be decomposed as follows:

$$\underbrace{H^T - H^N}_{\Delta_H} = \underbrace{(C^T - C^N)(1 - D^T)}_{\Delta_{CH}} + \underbrace{C^N(D^N - D^T)}_{\Delta_D},$$

such that the change in the Human Opportunity Index can be decomposed in two components:  $\Delta_{CH}$ , the contribution of the change in average school re-enrollment, and  $\Delta_D$ , the contribution of the change in inequality of opportunity. After computing  $G^T$ ,  $S^T$ ,  $G^N$  and  $S^N$ , following the decomposition procedure described in (4.2), the difference in the Gini Opportunity Aggregator can be decomposed similarly:

$$\underbrace{S^T - S^N}_{\Delta_S} = \underbrace{(C^T - C^N)(1 - G^T)}_{\Delta_{CG}} + \underbrace{C^N(G^N - G^T)}_{\Delta_G}$$

where we see that the change in the Gini Opportunity Aggregator can be decomposed in two components:  $\Delta_{CG}$ , the contribution of the change in average school re-enrollment, and  $\Delta_G$ , the contribution of the change in inequality of opportunity.

## 4.4 Empirical Results

We estimate a logistic regression for each grade, where we estimate the probability that a child is re-enrolled as a function of his circumstances (the regression coefficients are reported in Table A.2 in the Appendix). The results suggest that treated children, boys and indigenous children have a higher probability of re-enrollment. Parental education plays an important role: living in a household where the head has a higher level of education (a proxy we use for education of the father) and, especially where the wife of the household head has a higher level of education (a proxy for maternal education) both correlate with a higher probability of school re-enrollment. Living in a locality where a secondary school is present also correlates positively with school re-enrollment, especially after completion of primary school (grade 6). The only surprising result is that indigenous children have a higher probability of re-enrollment, but remember that we control for other features of the household in the estimation, such as the level of education of the household head and his wife. Some of the popular impression that indigenous children have lower rates of re-enrollment than non-indigenous children might be due to the fact that they live in households with lower educated parents. No firm conclusions can be drawn about the interaction effects between treatment and circumstances: only very few of them are statistically significant<sup>15</sup>. It is striking, though, that, when the interaction effects are statistically significant, they usually have the opposite sign of the direct effect of the circumstance. This suggests that the treatment might compensate the effect of the circumstance on school re-enrollment, which can be expected to diminish inequality of opportunity.

### 4.4.1 Average effect

We use the estimated coefficients to predict for each child the probability of re-enrollment in case he is treated ( $p_i^T$ ), and in case he is not treated ( $p_i^N$ )<sup>16</sup>. We compare the average of these two probabilities per grade group and determine which percentage of the children (would) gain from receiving Progresá benefits (i.e., for which  $p_i^T > p_i^N$ ). The results are reported in Table 1.

The first column (I) gives the grade the children completed before the start of the school year. The second column (II) gives the sample sizes of the treatment (TS) and the third

<sup>15</sup>We don't discuss the coefficients of the missing value dummies, as we only included these dummies to keep our sample as large as possible –see also final comment in the previous Section.

<sup>16</sup>In the regressions some variables perfectly predicted school enrollment. They were then left out of the estimation (see Table A.2). In that case we deleted one of the co-linear variables, resulting in empty cells in Table A.2. See footnote in the table for further details



**Table 4.1:** Samples sizes and average (predicted) re-enrollment rates per grade in 1998

(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
Grade	Sample size		Av. Re-enroll.		Av. $p_i^T$		Av. $p_i^N$		% gains ( $p_t > p_n$ )
	TS	CS	TS	CS	TS	CS	TS	CS	TS $\cup$ CS
0	2803	1773	0.786	0.774	0.786	0.776	0.780	0.774	0.522
1	2139	1333	0.973	0.947	0.973	0.973	0.947	0.947	0.808
2	2116	1316	0.943	0.920	0.943	0.940	0.923	0.920	0.735
3	2011	1202	0.928	0.889	0.928	0.926	0.890	0.889	0.850
4	1773	1078	0.925	0.880	0.925	0.923	0.884	0.880	0.749
5	1640	979	0.914	0.870	0.914	0.913	0.871	0.870	0.922
6	2555	1519	0.520	0.397	0.520	0.517	0.400	0.397	0.994
7	803	474	0.929	0.930	0.929	0.929	0.929	0.930	0.534
8	645	345	0.912	0.890	0.912	0.912	0.883	0.890	0.758
9	449	254	0.454	0.433	0.454	0.462	0.438	0.433	0.560

Notes: TS(CS) stand for the sample of children that lived in a treatment (control) locality. TSUCN indicates that the union of both samples is used. Source: Author's calculations.

(III) that of the control sample (CS). The fourth and fifth columns show the re-enrollment rates for these samples; these numbers are the same as those depicted in Figure 1. Columns (VI)-(IX) give the predicted average probabilities of re-enrollment in case the child receives treatment ( $p^T$ ) and in case he does not receive treatment ( $p^N$ ), for the children in both the treatment and control sample. The final column, (X), gives the percentage of children for which the predicted probability in case of treatment is larger than the predicted probability of enrollment without treatment.

The following observations can be made. First, the numbers in column (IV) and (VI) are equal, as are the numbers in columns (V) and (IX). This is a result of the estimation procedure we followed. Second, the average probabilities in case of receiving treatment for the TS and CS at the one hand (columns VI and VII) and the average probabilities in case of not receiving treatment for the TS and CS on the other hand (columns VIII and IX), are very close to each other. This suggests that differences in re-enrollment rates between the two samples are not due to differences in composition in terms of the circumstances we incorporated. Third, differences between the average predicted re-enrollment probabilities  $\hat{p}^T$  and  $\hat{p}^N$  for both TS (column VI versus VIII) and CS (column VII versus IX) are very small for grades 0 and 7. Except for grade 5, the differences from grade 1 onwards become more pronounced, and they are particularly large for grade 6, i.e., for those that completed primary education. This suggests that Progresá induces children to stay in school, especially after completion of primary education. Finally, the last column confirms that for most children the predicted probability of re-enrollment is larger when they receive treatment than when

they don't. The percentage that gains is very large for children that completed grades 1 to 6. So, overall, more children seem to gain from the program than lose. Yet, the question what the program does to the evaluation of aggregated opportunities when inequality of opportunity is taken into account, or what the program does to inequality of opportunities remains unanswered so far.

#### 4.4.2 Dominance

We discuss in first place the results for Generalized Lorenz dominance. As the results for grade 6 are more outspoken and this is the grade where most children drop from school, we depict in Figure 2 the Generalized Lorenz Curve for this grade only. Results for the other grades are presented in Appendix 4.C, in Table A.4. As can be observed in Figure 4.2, the curve for treated children lies, for all deciles, above the one for non-treated children, suggesting that Progresá improves aggregate opportunities for the participants. That is, for all opportunity aggregator functions satisfying the three properties mentioned in Proposition 1, the distribution of opportunities in case of treatment is unambiguously preferred to the distribution in the absence of treatment. Furthermore, the difference between both curves is substantial and statistically significant.

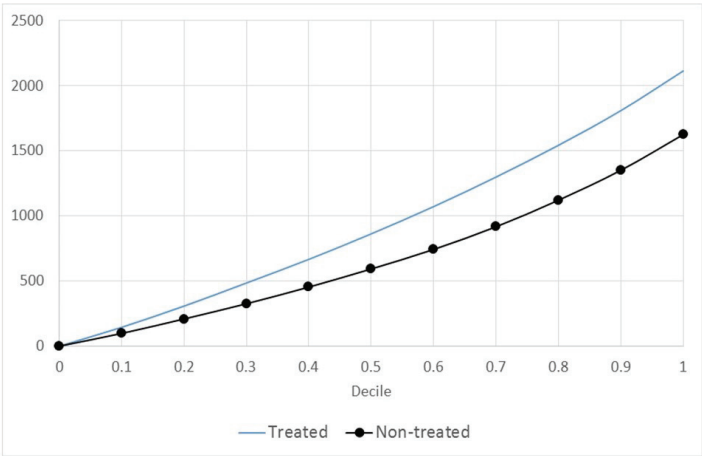
We now use Proposition 2 to look into the effects of the program on inequality of opportunity; we compare the Lorenz curve when children are treated and not treated. Again, the results for grade 6 are more interesting, so we show only results for this grade in Figure 4.3, but the reader can find in Table A.8. (Appendix 4.D) the coordinates for all the grades. Looking at the figure, it is clear that the Lorenz curve when treated is never below when not treated. Moreover, at each decile, the difference of the value of the coordinates of the Lorenz Curve between  $M_d^T$  and  $M_d^N$  is always statistically significant. Hence we can infer that Progresá reduced inequality of opportunity in grade 6 for all inequality measures satisfying the three properties mentioned in Proposition 2.

The results for the other grades are, overall also positive. For most grades, the differences ( $M_d^D$  and  $L_d^D$ ) are always positive, suggesting that the program reduces inequality of opportunities for those grades too. Clearly, for grades 0, 5 and 7, some values for  $M_d^D$  or  $L_d^D$  are negative, but these negative differences are small and never statistically significant.

#### 4.4.3 Decomposing the effect

The final step in our analysis is to decompose the effect of Progresá on our two selected aggregator functions, the Human Opportunity Index and the Gini Opportunity Aggregator, into an effect

Figure 4.2: Generalized Lorenz Curve grade 6



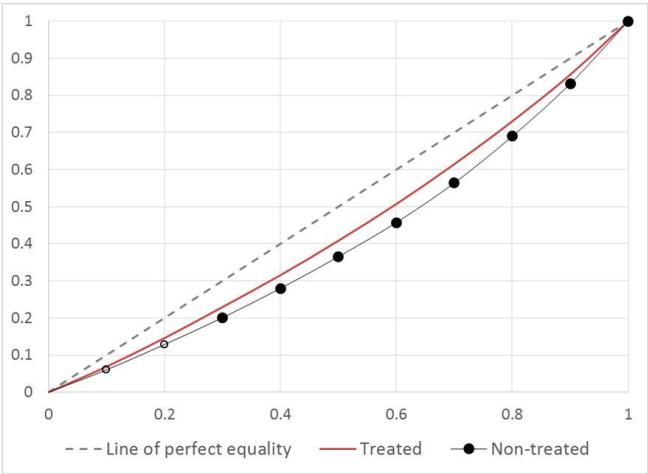
Notes: A solid-black circle (●) indicates the difference between treated and non-treated Generalized Lorenz curves, at a particular decile, is significant at 1% level of significance, while a white-unfilled circle (○) indicates that this difference is significant at 5%.

on average opportunities and inequality of opportunity. In Table 4.2 we report the results for the Human Opportunity Index.

The first column in the table gives the grade attained before the start of the school year. Column (II) and (III) give the value of the Human Opportunity Index based on the estimated probability of re-enrollment when treated ( $p_i^T$ ) and when non-treated ( $p_i^N$ ), respectively. Similarly, columns (V) and (VI) show the average re-enrollment rate, and columns (VIII) and (IX) the dissimilarity index. Column (IV) gives the change in the Human Opportunity Index, and equals the difference between column (II) and (III). Similarly, columns (VII) and (X) give the change in the average re-enrollment rate and dissimilarity index and equal the difference between columns (V) and (VI), and (VIII) and (IX), respectively. Column (XI) gives the part of the change in the Human Opportunity Index that can be attributed to an increase in average re-enrollment, and column (XII) that of the change that can be attributed to changes in the dissimilarity index. As the values of all level statistics are determined very precisely (columns II, III, V, VI, VIII, and IX), we only report whether the differences are statistically significant (columns IV, VII, IX and XII).

The following conclusions can be drawn. First, the values of the Human Opportunity

Figure 4.3: Lorenz Curve grade 6



Notes: A solid-black circle (●) indicates the difference between treated and non-treated Lorenz curves, at a particular decile, is significant at 1% level of significance, while a white-unfilled circle (○) indicates that this difference is significant at 5%.

Index are very high for all grades, except for grades 6 and 9, irrespective of whether children receive treatment or not. Second, there exist gains from treatment in terms of the Human Opportunity Index for almost all grades, the exceptions being grades 5 and 7, but these losses are rather small and not statistically significant (column IV). Third, the positive effect observed in grades 2, 3, 4, and 6, is due to an increase in average re-enrollment (column VII), and to a decrease in the dissimilarity index (column X). Finally, the contribution of the decrease in inequality of opportunity,  $\Delta_D$ , to the improvement of the Human Opportunity Index in those grades in it increased significantly at 5% (i.e., grades 2, 3, 4 and 6), accounts for between 16% and 32% of the improvement in the value of the Human Opportunity Index.

Table 4.3 reports the results for the Gini Opportunity Aggregator ( $S$ ), using the same format as in Table 4.2, but without repeating the results for average re-enrollment, as they are identical to the ones reported in this table. The results are entirely in line with those reported in and discussed after Table 2. The only difference is that the effects on the Gini Opportunity Aggregator are usually slightly larger. Clearly, the results tell the same story: Progresá not only improved average re-enrollment rates of children, but it also decreased inequality of opportunity. The effects are statistically significant for grades 2, 3, 4 and 6, and

**Table 4.2:** The Human Opportunity Index (all entries multiplied by 100)

(I) Grade	(II) Human T	(III) Opp. N	(IV) Index T-N	(V) Re-enrollment T	(VI) rate N	(VII) T-N	(VIII) Dissimilarity T	(IX) Index C	(X) T-N	(XI) $\Delta_{CH}$	(XII) $\Delta_D$
0	74.025	73.701	0.324	78.215	77.772	0.444	5.357	5.233	0.123	0.420	-0.096
1	96.472	92.723	3.749	97.319	94.218	3.102	0.871	1.587	-0.715	3.075	0.674
2	92.500	90.163	2.337 **	94.175	92.235	1.941 **	1.779	2.246	-0.467	1.906 **	0.430
3	91.388	86.950	4.438 ***	92.733	88.970	3.762 ***	1.450	2.270	-0.821 **	3.708 ***	0.730 **
4	91.273	85.279	5.993 ***	92.433	88.298	4.135 ***	1.255	3.418	-2.163 ***	4.083 ***	1.910 ***
5	89.288	89.434	-0.146	91.360	91.043	0.317	2.268	1.768	0.500	0.309	-0.456
6	47.034	34.210	12.824 ***	51.903	39.910	11.992 ***	9.380	14.283	-4.903 ***	10.867 ***	1.957 ***
7	91.415	91.645	-0.229	92.899	92.967	-0.067	1.597	1.422	0.175	-0.066	-0.163
8	89.538	86.094	3.445 *	91.195	88.581	2.614	1.816	2.807	-0.991	2.567	0.878
9	41.396	37.393	4.003	45.723	43.656	2.067	9.464	14.347	-4.883	1.871	2.132

Notes: \*denotes significance at 10%, \*\* at 5%, and \*\*\* at 1%. Confidence intervals were determined using a Bootstrap procedure. Source: Author’s calculations.

are especially large for those children that completed primary education (grade 6) and grade 4. The contribution of the decrease in inequality of opportunity,  $\Delta_G$ , to the improvement of the Gini Opportunity Aggregator in those grades in which it increased significantly at 5% (i.e., grades 2, 3, 4 and 6), accounts for between 19% and 39% of the improvement in the value of the Gini Opportunity Aggregator.

4.5 Conclusion

Many social programs aim to improve children’s opportunities, especially for those that, due to their circumstances, would not get sufficient opportunities. From this perspective, it is surprising that most program evaluation studies only identify an average treatment effect. Some studies focus on children with particular circumstances (girls versus boys or indigenous versus non-indigenous), and identify for which types of children the program works better. However, so far, no study has tried to obtain an overall assessment of the effect of a social program like Progresa on the distribution of school re-enrollment between children with different circumstances. This paper is a first attempt to provide an assessment of the effects of

**Table 4.3:** The Gini Social Welfare Function ( $S$ ) (all entries multiplied by 100)

(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)
Grade	T	Gini Welfare N	T-N	T	Gini N	T-N	$\Delta_{CG}$	$\Delta_G$
0	72.381	72.062	0.319	7.459	7.341	0.117	0.410	-0.091
1	96.175	92.115	4.059	1.176	2.231	-1.055	3.065	0.994
2	91.957	89.336	2.621 **	2.355	3.143	-0.787	1.895 **	0.726 *
3	90.840	86.087	4.752 ***	2.041	3.240	-1.199 **	3.685 ***	1.067 **
4	90.778	84.121	6.657 ***	1.790	4.730	-2.939 ***	4.061 ***	2.595 ***
5	88.522	88.899	-0.376	3.106	2.355	0.750	0.306	-0.683
6	45.332	32.439	12.892 ***	12.660	18.721	-6.061 ***	10.473 ***	2.419 ***
7	90.934	91.132	-0.1975	2.115	1.974	0.141	-0.065	-0.131
8	88.846	85.097	3.748 *	2.575	3.932	-1.356	2.546	1.202
9	39.687	35.041	4.646	13.201	19.735	-6.534	1.794	2.852

*Notes: \*denotes significance at 10%, \*\* at 5%, and \*\*\* at 1%. Confidence intervals were determined using a Bootstrap procedure. Source: Author's calculations.*

Progresá that takes into account its effect on the distribution of children's opportunities.

To do so, we find inspiration in the recent literature that tries to quantify the extent to which opportunities are unequally distributed. A child's probability of being re-enrolled, conditional on its circumstances is seen as a measure of the opportunities available to the child. Inequalities in these probabilities are exclusively due to children having different circumstances and are therefore considered offensive. Hence a social planner that wants to evaluate the distribution of opportunities should favor a redistribution (i.e., a transfer) of opportunities from those with higher to those with lower opportunities. The evaluation framework we develop here incorporates this concern.

We have seen that, overall, the distribution of re-enrollment probabilities conditional on

circumstances when treated both Generalized Lorenz and Lorenz dominates the distribution of re-enrollment probabilities when not treated. The former indicates that any social evaluator that aggregates children's opportunities and is inequality averse with respect to the distribution of opportunities must find that Progresa increases aggregate opportunities. The latter indicates that, whenever inequality of opportunity is measured by a relative measure, Progresa decreases inequality of opportunities. Finally, when the Human Opportunity Index or the Gini Opportunity Aggregator is chosen to aggregate children's opportunities, the effect on the reduction in inequality of opportunity by the program accounts for between 15 and 40% of the effect of the program on the aggregate opportunities. Hence, we found clear evidence that Progresa not only improved school re-enrollment opportunities on average, but it also significantly and substantially reduced inequality of school re-enrollment opportunities.

## Bibliography

- ATTANASIO, O. P., C. MEGHIR, AND A. SANTIAGO (2012): "Education choices in Mexico: using a structural model and a randomized experiment to evaluate Progresa," *The Review of Economic Studies*, 79, 37–66.
- BEHRMAN, J., AND P. TODD (1999): "Randomness in the experimental samples of PROGRESA –Education, Health, and Nutrition Program," *International Food Policy Research Institute*.
- BEHRMAN, J. R., P. SENGUPTA, AND P. TODD (2005): "Progressing through PROGRESA: an impact assessment of a school subsidy experiment in rural Mexico," *Economic Development and Cultural Change*, 54, 237–275.
- COHEN, G. A. (1989): "On the currency of egalitarian justice," *Ethics*, pp. 906–944.
- DE BARROS, R., F. FERREIRA, J. MOLINAS VEGA, AND J. CHANDUVI (2009): *Measuring Inequality of Opportunities in Latin America and the Caribbean*. The World Bank.
- DE BARROS, R. P., J. R. M. VEGA, AND J. SAAVEDRA (2008): "Measuring inequality of opportunities for children. Unpublished.," *The World Bank*.
- DUBOIS, P., A. DE JANVRY, AND E. SADOULET (2012): "Effects on school enrollment and performance of a conditional cash transfer program in Mexico," *Journal of Labor Economics*, 30, 555–589.
- FERREIRA, F. H. G., AND J. GIGNOUX (2011): "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America," *Review of Income and Wealth*, forthcoming.
- FERREIRA H.G., F., AND V. PERAGINE (2015): "Equality of opportunity: Theory and evidence," *Policy Research Working paper, No 7217. The World Bank. Washington, D.C.*
- FIGUEROA, J. L. (2014): "Distributional effects of Oportunidades on early child development," *Social Science & Medicine*, 113, 42–49.
- FLEURBAEY, M. (2008): *Fairness, Responsibility, and Welfare*. Oxford University Press.
- FOSTER, J. E., AND A. F. SHORROCKS (1988): "Inequality and poverty orderings," *European Economic Review*, 32(2), 654–661.



- GERTLER, P. (2004): "Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment," *American Economic Review*, pp. 336–341.
- LAMBERT, P. J. (2001): *The Distribution and Redistribution of Income: Third edition*. Manchester University Press.
- MOLINAS, J. R., R. P. DE BARROS, J. SAAVEDRA, AND M. GUIGALE (2012): "Do our children have a chance? A Human Opportunity Report for Latin America and the Caribbean," *The World Bank, Washington, D.C.*
- PISTOLESI, N. (2009): "Inequality of opportunity in the land of opportunities, 1968–2001," *The Journal of Economic Inequality*, 7, 411–433.
- RAMOS, X., AND D. VAN DE GAER (2015): "Approaches to inequality of opportunity: Principles, measures, and evidence," *Journal of Economic Surveys*.
- ROEMER, J. (1993): "A Pragmatic Theory of Responsibility for the Egalitarian Planner," *Philosophy & Public Affairs*, 22(2), 146–166.
- (1998): *Equality of opportunity*. Harvard University Press, Cambridge MA.
- ROEMER, J. E., AND A. TRANNOY (2014): *Equality of Opportunity* chap. 4, pp. 217–300. North-Holland.
- SCHULTZ, P. (2004a): "School subsidies for the poor: evaluating the Mexican Progresa poverty program," *Journal of Development Economics*, 74(1), 199–250.
- SCHULTZ, T. P. (2004b): "School subsidies for the poor: evaluating the Mexican Progresa poverty program," *Journal of Development Economics*, 74, 199–250.
- SHORROCKS, A. F. (1983): "Ranking income distributions," *Economica*, 50, 3–17.
- SKOUFIAS, E. (2005): *PROGRESA and its impacts on the welfare of rural households in Mexico*, vol. 139. Intl Food Policy Res Inst.
- TODD, P. E., AND K. I. WOLPIN (2006): "Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility," *The American Economic Review*, pp. 1384–1417.
- VAN DE GAER, D., J. VANDENBOSSCHE, AND J. L. FIGUEROA (2014): "Children's health opportunities and project evaluation: Mexico's Oportunidades program," *The World Bank Economic Review*, 28, 282–310.



# Appendices

## 4.A Composition of the sample

**Table A.1:** Composition of and missing values in treatment and control samples in percentage

	TS	NS	TUN
(a)Composition			
Gender child (male)	48.01	46.61	47.48
Indigenous background	28.84	30.32	29.4
Gender household head (male)	86.71	87.18	86.89
Household head living with partner	90.72	90.63	90.69
Secondary school in the locality	29.46	28.61	29.14
Household head no education	25.47	26.48	25.85
Household head incomplete primary	45.83	45.45	45.69
Household head at least complete primary	17.26	15.9	16.74
Spouse no education	28	29.98	28.37
Spouse incomplete primary	38.54	36.3	37.69
Spouse at least complete primary	14.4	13.76	14.16
(b) Missing values			
Gender child (male)	7.67	8.07	7.82
Indigenous background	14.99	16.25	15.47
Gender household head	5.97	5.59	5.83
Household head living with partner	1.2	0.73	1.02
Secondary school in the locality	0.57	0	0.36
Education of household head	11.44	12.18	11.72
Education of spouse	19.07	20.96	19.79

*Notes: TS(NS) stands for the sample of children that lived in a treatment(control) locality; TS $\cup$ CN indicates that the union of both samples is used In part (a), columns 2-4 give the fraction of the children that have the characteristics mentioned in the first column, as a percentage of the children in that sample for which data on that characteristic were reported In part (b), columns 2-4 give the fraction of children for which the characteristic mentioned in the first column is missing, as a percentage of all the children in the sample. Source: Author's calculations.*

Table A.2: Estimated coefficients of the logistic model for school enrollment per grade

	Grade in 1998									
	0	1	2	3	4	5	6	7	8	9
T	0.554* (0.312)	-0.0979 (0.755)	-0.196 (0.429)	0.156 (0.507)	1.607*** (0.579)	0.417 (0.482)	0.719*** (0.259)	-0.350 (0.715)	0.112 (0.811)	-0.231 (0.686)
Ch sex	0.246** (0.106)	-0.351 (0.289)	0.228 (0.206)	-0.213 (0.208)	0.173 (0.208)	0.0798 (0.195)	0.294*** (0.112)	0.435 (0.345)	-0.192 (0.363)	0.582** (0.268)
Ind	0.0468 (0.200)	0.0668 (0.308)	-0.182 (0.264)	0.494* (0.292)	0.255 (0.236)	0.663** (0.295)	0.367** (0.165)	0.551 (0.456)	0.582 (0.497)	0.694* (0.356)
Hh sex	-0.125 (0.247)	-0.421 (0.547)	-0.755** (0.363)	-0.428 (0.385)	-0.132 (0.446)	-0.0917 (0.290)	0.131 (0.183)	0.610 (0.649)	0.487 (0.863)	0.297 (0.593)
Hh married	0.451* (0.256)	0.838* (0.487)	0.505 (0.395)	0.627* (0.368)	0.590 (0.378)	-0.100 (0.378)	0.0185 (0.207)	-0.628 (0.790)	-0.0668 (1.006)	-0.408 (0.600)
Secondary loc	0.166 (0.219)	0.170 (0.413)	0.371 (0.243)	0.0204 (0.210)	0.556** (0.248)	0.394 (0.363)	1.055*** (0.143)	0.619 (0.432)	-0.357 (0.393)	-0.825** (0.348)
Hh Inc prim	0.705*** (0.169)	0.410 (0.455)	0.618** (0.276)	0.234 (0.214)	-0.0433 (0.257)	0.143 (0.229)	0.0766 (0.138)	-0.277 (0.406)	-1.024** (0.443)	-0.371 (0.389)
Hh Comp prim	0.709*** (0.217)	1.085* (0.577)	0.676 (0.461)	0.772** (0.356)	1.156** (0.466)	0.381 (0.385)	0.632*** (0.228)	-0.142 (0.678)	-1.508** (0.588)	-0.321 (0.390)
Sp Inc prim	0.220 (0.213)	0.514 (0.371)	0.525** (0.227)	0.234 (0.232)	0.781** (0.319)	0.0429 (0.208)	0.117 (0.163)	-0.00734 (0.417)	0.586 (0.410)	0.629* (0.380)
Sp Comp prim	1.231*** (0.326)	1.975*** (0.741)	2.978*** (0.957)	1.340*** (0.504)	1.637*** (0.574)	1.219** (0.529)	0.520** (0.222)	0.634 (0.698)	1.568** (0.661)	0.434 (0.464)
Ch Sex * T	-0.178 (0.139)	1.167*** (0.384)	-0.0556 (0.270)	0.610** (0.250)	-0.285 (0.273)	-0.121 (0.307)	-0.103 (0.146)	-0.185 (0.440)	0.0267 (0.480)	-0.361 (0.322)
Ind * T	-0.394* (0.198)	0.391 (0.384)	0.0576 (0.270)	-0.451 (0.250)	0.0651 (0.273)	0.0323 (0.307)	0.0215 (0.146)	-0.768 (0.440)	0.402 (0.480)	-0.103 (0.322)

Continued on next page

Table A.2 – continued from previous page

Grade in 1998										
	0	1	2	3	4	5	6	7	8	9
Hh Sex * T	(0.226) 0.130 (0.308)	(0.453) 0.123 (0.663)	(0.376) 0.282 (0.460)	(0.351) 0.379 (0.569)	(0.387) -0.0955 (0.574)	(0.406) -0.426 (0.430)	(0.224) -0.465** (0.227)	(0.523) -0.600 (0.767)	(0.662) -0.627 (0.896)	(0.464) -0.668 (0.680)
Hh married * T	(0.316) -0.514 (0.665)	(0.665) -0.0379 (0.665)	(0.549) 0.551 (0.549)	(0.518) -0.134 (0.518)	(0.514) -0.898* (0.514)	(0.494) 0.187 (0.494)	(0.255) 0.251 (0.255)	(0.922) 0.670 (0.922)	(1.043) 0.345 (1.043)	(0.696) 0.813 (0.696)
Secondary loc * T	(0.274) 0.116 (0.274)	(0.539) 0.548 (0.539)	(0.388) -0.312 (0.388)	(0.302) -0.125 (0.302)	(0.378) -0.206 (0.378)	(0.458) 0.0875 (0.458)	(0.214) -0.293 (0.214)	(0.529) -0.500 (0.529)	(0.468) 0.381 (0.468)	(0.408) 0.668 (0.408)
Hh inc prim * T	(0.223) -0.222 (0.223)	(0.558) 0.314 (0.558)	(0.388) -0.398 (0.388)	(0.265) 0.107 (0.265)	(0.322) 0.243 (0.322)	(0.303) -0.0983 (0.303)	(0.169) 0.106 (0.169)	(0.517) 0.784 (0.517)	(0.523) 0.515 (0.523)	(0.464) 0.212 (0.464)
Hh comp prim * T	(0.312) 0.586* (0.312)	(0.743) -0.274 (0.743)	(0.750) 0.511 (0.750)	(0.495) 0.350 (0.495)	(0.636) -0.462 (0.636)	(0.441) 0.214 (0.441)	(0.285) -0.0645 (0.285)	(0.836) 0.825 (0.836)	(0.727) 0.871 (0.727)	(0.486) 0.671 (0.486)
Sp inc prim * T	(0.158) (0.226)	(0.495) (0.495)	(0.385) (0.385)	(0.273) (0.273)	(0.411) (0.411)	(0.326) (0.326)	(0.184) (0.184)	(0.543) (0.543)	(0.554) (0.554)	(0.439) (0.439)
Sp comp prim * T	(0.417 (0.387)	(0.571 (0.940)	(-1.562 (1.122)	(-0.565 (0.659)	(-1.046 (0.701)	(0.112 (0.666)	(0.211 (0.258)	(0.277 (0.904)	(-1.733** (0.825)	(-0.129 (0.583)
Ch sex m	-0.298* (0.154)			0.239 (0.278)					0.708 (0.582)	
Hh educ m	0.491*** (0.166)		1.141*** (0.400)	0.610** (0.279)						
Sp educ m		0.932*** (0.283)							0.285 (0.481)	
Secondary loc m	0.457*** (0.0982)									
Hh married m	1.332***				-1.313**					

Continued on next page

Continued on next page

Table A.2 – continued from previous page

Grade in 1998									
	0	1	2	3	4	5	6	7	8
Hh sex m	(0.410)		-1.506*** (0.533)		(0.549)				-1.187** (0.560)
Ind m									-0.534 (0.529)
Constant	0.155 (0.263)	1.797*** (0.567)	1.818*** (0.427)	1.388*** (0.379)	0.879** (0.368)	1.538*** (0.303)	-1.265*** (0.203)	2.176*** (0.649)	2.252*** (0.862)
Observations	4,576	3,472	3,432	3,213	2,851	2,619	4,074	1,277	990
									703

Notes: a) Robust standard errors in parentheses. The stars behind the coefficients indicate their significance: \*\*\* (\*\*) [\*] means significant at 1 (5) [10] percent respectively.

- b) Empty cells indicate either that the corresponding dependent variable was perfectly collinear with (a subset of) the other variables or that the coefficient was not significant in a previous specification (one that included a dummy for each missing characteristic).
- c) Keeping only missing dummies that are significant (in the original specification) avoids dependent variables that perfectly predict school enrollment. This point turned particularly important for the bootstrap procedure described in Appendix 4.B because perfectly predicted variables impeded achieving convergence. That is the reason that in some cases variables that were originally significant were not any more in the final specification, as is the case for “Ch sex m” in grade 3.
- d) All variables in the regression are dummy variables: they are either 1 or zero. We indicate when they are equal to 1. Variable T: equal to 1 if the child is treated; Ch sex: equal to 1 if the child is a boy; Ind: equal to 1 for indigenous children; Hh sex: equal to 1 if the household head is male; Hh married: equal to 1 if the household head is married; Secondary loc: equal to 1 if there is a Secondary school in the locality where the child lives; Hh Inc prim (Sp Inc prim): equal to 1 if the (spouse of the) household head has some schooling, but did not complete primary education; Hh Comp prim (Sp Comp prim) : equal to 1 if the (spouse of the) household head completed at least primary schooling; Ch Sex m: equal to 1 if the child’s gender is missing; Ind m: equal to 1 if the indigenous background of the child is missing; Hh Sex m: equal to 1 if the household head’s gender is missing; Hh married m: equal to 1 if the household head’s marital status is missing; Secondary loc m: equal

to 1 if we don't know whether there is a secondary school in the locality where the child lives; Hh educ m (Sp educ m): equal to 1 if the (spouse of the) household head's education is missing. Source: Authors' calculations.

Table A.3: Preprogram characteristics of the sample

	Grade 0				Grade 1				Grade 2			
	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t
Ch sex	0.469	0.466	0.003	0.158	0.483	0.479	0.004	0.2	0.482	0.473	0.009	0.462
Ind	0.291	0.304	-0.013	-0.285	0.29	0.294	-0.004	-0.089	0.284	0.298	-0.014	-0.293
Hh sex	0.858	0.882	-0.024	-2.13	0.875	0.869	0.006	0.382	0.866	0.875	-0.009	-0.586
Hh married	0.909	0.909	0	0.019	0.917	0.915	0.002	0.173	0.915	0.916	-0.001	-0.084
Secondary Loc	0.284	0.276	0.008	0.114	0.291	0.293	-0.002	-0.03	0.277	0.281	-0.004	-0.07
Hh Inc Prim	0.406	0.406	0	0.004	0.436	0.431	0.005	0.222	0.458	0.448	0.01	0.451
Hh Comp Prim	0.191	0.17	0.021	0.953	0.193	0.178	0.015	0.765	0.176	0.161	0.015	0.826
Sp inc prim	0.335	0.327	0.008	0.313	0.373	0.335	0.038	1.558	0.379	0.332	0.047	2.056
Sp comp prim	0.158	0.151	0.007	0.416	0.173	0.17	0.003	0.172	0.156	0.146	0.01	0.534
Ind m	0.225	0.222	0.003	0.163	0.159	0.165	-0.006	-0.369	0.133	0.156	-0.023	-1.8
Sp educ m	0.201	0.204	-0.003	-0.155	0.184	0.207	-0.023	-1.378	0.179	0.216	-0.037	-2.347
Hh educ m	0.131	0.127	0.004	0.257	0.122	0.122	0	0.034	0.112	0.126	-0.014	-1.198
Secondary Loc m	0.006	0	0.006	1.38	0.006	0	0.006	1.164	0.006	0	0.006	1.375
Hh married m	0.014	0.007	0.007	1.98	0.014	0.007	0.007	2.014	0.014	0.005	0.009	2.458
Hh sex m	0.067	0.054	0.013	1.504	0.059	0.053	0.006	0.781	0.061	0.059	0.002	0.205
dirtfloor	0.706	0.695	0.011	0.332	0.664	0.678	-0.014	-0.423	0.677	0.668	0.009	0.291
poor roof	0.737	0.712	0.025	0.728	0.718	0.724	-0.006	-0.167	0.73	0.701	0.029	0.927
poor wall	0.152	0.215	-0.063	-2.019	0.138	0.206	-0.068	-2.263	0.142	0.183	-0.041	-1.31
water land	0.295	0.243	0.052	1.122	0.314	0.25	0.064	1.325	0.328	0.255	0.073	1.465
water house	0.046	0.036	0.01	1.007	0.053	0.04	0.013	1.234	0.051	0.041	0.01	0.848
toilet	0.469	0.462	0.007	0.188	0.501	0.482	0.019	0.533	0.513	0.511	0.002	0.038
electricity	0.568	0.596	-0.028	-0.619	0.596	0.614	-0.018	-0.435	0.611	0.634	-0.023	-0.573
blender	0.143	0.187	-0.044	-2.034	0.176	0.205	-0.029	-1.153	0.184	0.217	-0.033	-1.362
fridge	0.037	0.033	0.004	0.325	0.053	0.045	0.008	0.607	0.052	0.051	0.001	0.114
gas stove	0.149	0.154	-0.005	-0.183	0.163	0.185	-0.022	-0.806	0.173	0.184	-0.011	-0.391

Continued on next page



Table A.3 – continued from previous page

	Grade 0				Grade 1				Grade 2			
	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t
gas heater	0.016	0.014	0.002	0.481	0.018	0.017	0.001	0.087	0.021	0.011	0.01	1.77
radio	0.475	0.514	-0.039	-1.216	0.534	0.528	0.006	0.212	0.535	0.537	-0.002	-0.062
dvd	0.027	0.03	-0.003	-0.376	0.037	0.032	0.005	0.825	0.033	0.025	0.008	1.107
tv	0.281	0.309	-0.028	-0.958	0.334	0.364	-0.03	-0.953	0.315	0.382	-0.067	-2.112
video	0.014	0.012	0.002	0.415	0.015	0.013	0.002	0.477	0.012	0.014	-0.002	-0.358
wash machine	0.011	0.008	0.003	0.737	0.018	0.011	0.007	1.033	0.02	0.016	0.004	0.588
fan	0.021	0.038	-0.017	-1.768	0.029	0.038	-0.009	-0.878	0.028	0.039	-0.011	-1.063
car	0.003	0.002	0.001	0.792	0.007	0.002	0.005	1.888	0.007	0.004	0.003	0.932
truck	0.012	0.012	0	-0.011	0.021	0.023	-0.002	-0.258	0.016	0.021	-0.005	-0.702
land ownership	0.577	0.547	0.03	0.766	0.586	0.582	0.004	0.106	0.608	0.586	0.022	0.667
animals ownership	0.35	0.321	0.029	0.948	0.381	0.331	0.05	1.633	0.38	0.353	0.027	0.838
workhead	0.752	0.787	-0.035	-1.448	0.772	0.768	0.004	0.185	0.786	0.786	0	0.01
workspouse	0.067	0.062	0.005	0.324	0.06	0.058	0.002	0.169	0.069	0.052	0.017	1.045
Household charac m	0.068	0.081	-0.013	-1.022	0.071	0.08	-0.009	-0.667	0.055	0.071	-0.016	-1.449
assets m	0.713	0.688	0.025	0.767	0.726	0.71	0.016	0.546	0.73	0.713	0.017	0.607
demographaphics m	0.058	0.068	-0.01	-0.796	0.057	0.067	-0.01	-0.796	0.045	0.061	-0.016	-1.512
N rooms	1.46	1.454	0.006	0.119	1.512	1.48	0.032	0.701	1.571	1.523	0.048	0.957
Hh age	34.67	35.912	-1.242	-1.505	35.293	36.116	-0.823	-1.135	36.628	36.663	-0.035	-0.051
Spouse age	27.633	28.438	-0.805	-1.287	28.887	28.399	0.488	0.71	29.773	28.378	1.395	2.046
N children 0-5	1.565	1.526	0.039	0.923	1.283	1.201	0.082	1.607	1.303	1.223	0.08	1.612
N children 6-12	1.703	1.715	-0.012	-0.219	2.107	2.082	0.025	0.483	2.16	2.153	0.007	0.117
N children 13-15	0.539	0.541	-0.002	-0.082	0.529	0.512	0.017	0.624	0.591	0.6	-0.009	-0.356
N children 16-20	0.41	0.433	-0.023	-0.706	0.388	0.408	-0.02	-0.78	0.456	0.428	0.028	1.046
N women 20-39	0.758	0.747	0.011	0.591	0.757	0.744	0.013	0.673	0.745	0.727	0.018	0.765
N women 40-59	0.224	0.263	-0.039	-2.823	0.244	0.245	-0.001	-0.106	0.263	0.241	0.022	1.236
N women 60	0.084	0.095	-0.011	-1.087	0.08	0.09	-0.01	-0.789	0.096	0.093	0.003	0.291

Continued on next page



Table A.4: Preprogram characteristics of the sample

	Grade 3				Grade 4				Grade 5			
	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t
Ch sex	0.486	0.482	0.004	0.226	0.472	0.471	0.001	0.049	0.5	0.449	0.051	2.295
Ind	0.291	0.312	-0.021	-0.444	0.309	0.333	-0.024	-0.483	0.291	0.303	-0.012	-0.242
Hh sex	0.88	0.886	-0.006	-0.503	0.861	0.865	-0.004	-0.356	0.878	0.866	0.012	0.896
Hh married	0.909	0.905	0.004	0.318	0.902	0.904	-0.002	-0.105	0.904	0.903	0.001	0.084
Secondary Loc	0.297	0.25	0.047	0.845	0.306	0.281	0.025	0.4	0.298	0.301	-0.003	-0.062
Hh Inc Prim	0.456	0.455	0.001	0.056	0.453	0.455	-0.002	-0.08	0.488	0.475	0.013	0.484
Hh Comp Prim	0.186	0.157	0.029	1.419	0.17	0.149	0.021	1.062	0.166	0.149	0.017	0.801
Sp inc prim	0.37	0.382	-0.012	-0.453	0.39	0.355	0.035	1.398	0.4	0.381	0.019	0.755
Sp comp prim	0.149	0.131	0.018	1.115	0.138	0.124	0.014	0.83	0.137	0.124	0.013	0.675
Ind m	0.128	0.159	-0.031	-2	0.131	0.155	-0.024	-1.467	0.124	0.123	0.001	0.078
Sp educ m	0.185	0.215	-0.03	-1.753	0.188	0.214	-0.026	-1.619	0.188	0.194	-0.006	-0.331
Hh educ m	0.098	0.123	-0.025	-1.921	0.121	0.14	-0.019	-1.351	0.102	0.105	-0.003	-0.201
Secondary Loc m	0.005	0	0.005	1.271	0.005	0	0.005	1.222	0.004	0	0.004	1.282
Hh married m	0.01	0.01	0	0.104	0.011	0.007	0.004	0.986	0.012	0.01	0.002	0.277
Hh sex m	0.048	0.049	-0.001	-0.178	0.067	0.061	0.006	0.57	0.054	0.06	-0.006	-0.598
dirtfloor	0.668	0.676	-0.008	-0.231	0.668	0.682	-0.014	-0.474	0.668	0.681	-0.013	-0.449
poor roof	0.728	0.715	0.013	0.367	0.72	0.735	-0.015	-0.497	0.738	0.739	-0.001	-0.02
poor wall	0.144	0.194	-0.05	-1.526	0.152	0.203	-0.051	-1.65	0.149	0.215	-0.066	-1.803
water land	0.33	0.263	0.067	1.351	0.321	0.242	0.079	1.7	0.326	0.269	0.057	1.117
water house	0.056	0.037	0.019	1.507	0.053	0.036	0.017	1.332	0.054	0.047	0.007	0.531
toilet	0.513	0.503	0.01	0.272	0.532	0.536	-0.004	-0.117	0.56	0.565	-0.005	-0.134
electricity	0.621	0.638	-0.017	-0.394	0.607	0.626	-0.019	-0.457	0.63	0.677	-0.047	-1.155
blender	0.182	0.224	-0.042	-1.551	0.175	0.223	-0.048	-1.994	0.181	0.251	-0.07	-2.555
fridge	0.051	0.068	-0.017	-1.079	0.052	0.042	0.01	0.77	0.055	0.054	0.001	0.095
gas stove	0.167	0.175	-0.008	-0.297	0.169	0.182	-0.013	-0.429	0.162	0.196	-0.034	-1.015

Continued on next page



Table A.4 – continued from previous page

	Grade 3				Grade 4				Grade 5			
	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t
N men 20-39	0.566	0.584	-0.018	-0.787	0.545	0.544	0.001	0.072	0.52	0.54	-0.02	-0.695
N men 40-59	0.396	0.367	0.029	1.251	0.4	0.377	0.023	1.096	0.438	0.418	0.02	0.869
N men 60	0.08	0.092	-0.012	-0.987	0.068	0.083	-0.015	-1.245	0.095	0.099	-0.004	-0.3

Table A.5: Preprogram characteristics of the sample

	Grade 6			Grade 7			Grade 8					
	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t
Ch sex	0.447	0.423	0.024	1.196	0.517	0.54	-0.023	-0.777	0.538	0.443	0.095	2.924
Ind	0.266	0.283	-0.017	-0.389	0.286	0.297	-0.011	-0.228	0.285	0.31	-0.025	-0.471
Hh sex	0.863	0.854	0.009	0.663	0.864	0.873	-0.009	-0.425	0.853	0.867	-0.014	-0.584
Hh married	0.899	0.886	0.013	1.228	0.902	0.911	-0.009	-0.539	0.904	0.919	-0.015	-0.764
Secondary Loc	0.25	0.244	0.006	0.125	0.375	0.384	-0.009	-0.134	0.378	0.371	0.007	0.102
Hh Inc Prim	0.498	0.502	-0.004	-0.141	0.463	0.479	-0.016	-0.526	0.49	0.475	0.015	0.366
Hh Comp Prim	0.123	0.117	0.006	0.427	0.174	0.181	-0.007	-0.221	0.183	0.188	-0.005	-0.175
Sp inc prim	0.419	0.404	0.015	0.537	0.416	0.382	0.034	0.969	0.459	0.383	0.076	2.14
Sp comp prim	0.101	0.098	0.003	0.195	0.147	0.15	-0.003	-0.116	0.133	0.157	-0.024	-0.909
Ind m	0.132	0.149	-0.017	-1.417	0.147	0.135	0.012	0.498	0.112	0.148	-0.036	-1.632
Sp educ m	0.198	0.225	-0.027	-1.429	0.212	0.219	-0.007	-0.249	0.18	0.194	-0.014	-0.515
Hh educ m	0.106	0.122	-0.016	-1.174	0.121	0.108	0.013	0.666	0.112	0.11	0.002	0.064
Secondary Loc m	0.007	0	0.007	1.397	0.009	0	0.009	0.996	0.005	0	0.005	0.991
Hh married m	0.01	0.006	0.004	1.274	0.006	0.004	0.002	0.499	0.009	0.006	0.003	0.699
Hh sex m	0.058	0.062	-0.004	-0.49	0.052	0.049	0.003	0.307	0.071	0.058	0.013	0.839
dirtfloor	0.663	0.664	-0.001	-0.055	0.64	0.652	-0.012	-0.279	0.589	0.658	-0.069	-1.496
poor roof	0.71	0.729	-0.019	-0.545	0.704	0.717	-0.013	-0.318	0.727	0.725	0.002	0.057
poor wall	0.134	0.205	-0.071	-2.209	0.177	0.257	-0.08	-1.723	0.205	0.264	-0.059	-1.153
water land	0.333	0.255	0.078	1.725	0.366	0.283	0.083	1.707	0.398	0.252	0.146	2.537
water house	0.05	0.043	0.007	0.623	0.061	0.042	0.019	1.435	0.074	0.049	0.025	1.205
toilet	0.541	0.563	-0.022	-0.675	0.601	0.643	-0.042	-1.008	0.653	0.638	0.015	0.304
electricity	0.621	0.643	-0.022	-0.555	0.686	0.703	-0.017	-0.348	0.69	0.725	-0.035	-0.692
blender	0.192	0.25	-0.058	-2.07	0.23	0.295	-0.065	-2.131	0.245	0.31	-0.065	-1.856
fridge	0.049	0.053	-0.004	-0.34	0.062	0.061	0.001	0.049	0.073	0.061	0.012	0.579
gas stove	0.175	0.201	-0.026	-0.876	0.202	0.194	0.008	0.234	0.225	0.232	-0.007	-0.17

Continued on next page

Continued on next page

Table A.5 – continued from previous page

	Grade 6				Grade 7				Grade 8			
	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t
gas heater	0.019	0.016	0.003	0.616	0.02	0.011	0.009	1.358	0.019	0.014	0.005	0.586
radio	0.577	0.58	-0.003	-0.117	0.613	0.629	-0.016	-0.525	0.598	0.667	-0.069	-1.762
dvd	0.03	0.03	0	-0.078	0.042	0.017	0.025	2.57	0.047	0.023	0.024	1.941
tv	0.344	0.392	-0.048	-1.457	0.394	0.466	-0.072	-2.007	0.431	0.458	-0.027	-0.623
video	0.016	0.014	0.002	0.28	0.02	0.008	0.012	1.896	0.02	0.012	0.008	1.074
wash machine	0.014	0.019	-0.005	-0.688	0.034	0.021	0.013	1.084	0.028	0.035	-0.007	-0.474
fan	0.025	0.042	-0.017	-1.534	0.041	0.057	-0.016	-0.885	0.045	0.049	-0.004	-0.212
car	0.006	0.003	0.003	1.03	0.012	0.002	0.01	2.301	0.006	0	0.006	1.655
truck	0.02	0.029	-0.009	-0.794	0.024	0.025	-0.001	-0.148	0.028	0.023	0.005	0.43
land ownership	0.66	0.62	0.04	1.304	0.616	0.599	0.017	0.443	0.673	0.603	0.07	1.588
animals ownership	0.426	0.421	0.005	0.163	0.394	0.323	0.071	1.868	0.403	0.377	0.026	0.675
workhead	0.78	0.756	0.024	1.13	0.767	0.778	-0.011	-0.37	0.767	0.777	-0.01	-0.31
workspouse	0.047	0.057	-0.01	-0.98	0.051	0.036	0.015	1.117	0.064	0.038	0.026	1.573
Household charac m	0.055	0.072	-0.017	-1.436	0.08	0.059	0.021	1.192	0.054	0.055	-0.001	-0.048
assets m	0.787	0.759	0.028	1.26	0.758	0.724	0.034	0.944	0.797	0.733	0.064	1.795
demographics m	0.044	0.055	-0.011	-1.012	0.064	0.057	0.007	0.367	0.04	0.046	-0.006	-0.384
N rooms	1.681	1.672	0.009	0.143	1.687	1.675	0.012	0.189	1.78	1.736	0.044	0.656
Hh age	39.98	39.486	0.494	0.641	39.645	39.251	0.394	0.353	39.84	40.043	-0.203	-0.15
Spouse age	31.623	30.679	0.944	1.186	31.191	30.31	0.881	0.671	32.491	32.061	0.43	0.373
N children 0-5	0.968	0.966	0.002	0.055	0.917	0.876	0.041	0.595	0.839	0.872	-0.033	-0.528
N children 6-12	2.019	1.959	0.06	1.15	1.924	1.93	-0.006	-0.078	1.814	1.768	0.046	0.489
N children 13-15	0.978	1.001	-0.023	-0.781	1.014	0.937	0.077	1.345	1.208	1.168	0.04	0.829
N children 16-20	0.714	0.656	0.058	1.574	0.656	0.608	0.048	1.083	0.74	0.609	0.131	2.364
N women 20-39	0.648	0.633	0.015	0.67	0.624	0.656	-0.032	-0.901	0.681	0.664	0.017	0.411
N women 40-59	0.38	0.382	-0.002	-0.101	0.36	0.354	0.006	0.185	0.378	0.336	0.042	1.335
N women 60	0.078	0.096	-0.018	-1.357	0.093	0.108	-0.015	-0.655	0.101	0.116	-0.015	-0.676

Continued on next page

Table A.5 – continued from previous page

	Grade 6				Grade 7				Grade 8			
	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t	Treatment	Control	Diff	z/t
N men 20-39	0.47	0.51	-0.04	-1.549	0.427	0.485	-0.058	-1.676	0.456	0.478	-0.022	-0.533
N men 40-59	0.481	0.448	0.033	1.445	0.489	0.489	0	-0.001	0.516	0.475	0.041	1.18
N men 60	0.097	0.087	0.01	0.951	0.108	0.074	0.034	2.002	0.085	0.087	-0.002	-0.101



Table A.6: Preprogram characteristics of the sample

	Grade 9		
	Treatment	Control	Diff      z/t
Ch sex	0.508	0.48	0.028      0.686
Ind	0.312	0.319	-0.007    -0.118
Hh sex	0.86	0.894	-0.034    -1.36
Hh married	0.893	0.909	-0.016    -0.764
Secondary Loc	0.388	0.441	-0.053    -0.7
Hh Inc Prim	0.53	0.516	0.014    0.293
Hh Comp Prim	0.176	0.224	-0.048   -1.204
Sp inc prim	0.432	0.48	-0.048   -1.104
Sp comp prim	0.143	0.173	-0.03    -0.93
Ind m	0.138	0.122	0.016    0.669
Sp educ m	0.196	0.165	0.031    1.026
Hh educ m	0.125	0.087	0.038    1.691
Secondary Loc m	0.002	0	0.002    0.988
Hh married m	0.016	0.012	0.004    0.438
Hh sex m	0.065	0.039	0.026    1.465
dirtfloor	0.592	0.626	-0.034   -0.665
poor roof	0.735	0.685	0.05    0.953
poor wall	0.163	0.272	-0.109   -1.924
water land	0.41	0.35	0.06    0.848
water house	0.085	0.055	0.03    1.305
toilet	0.628	0.697	-0.069   -1.33
electricity	0.675	0.78	-0.105   -1.979
blender	0.252	0.311	-0.059   -1.578
fridge	0.082	0.047	0.035    1.48
gas stove	0.196	0.213	-0.017   -0.431

Continued on next page

Table A.6 – continued from previous page

	Grade 9			z/t
	Treatment	Control	Diff	
gas heater	0.016	0.02	-0.004	-0.292
radio	0.592	0.677	-0.085	-1.991
dvd	0.036	0.043	-0.007	-0.398
tv	0.412	0.488	-0.076	-1.496
video	0.031	0.02	0.011	1.215
wash machine	0.042	0.012	0.03	2.036
fan	0.065	0.055	0.01	0.336
car	0.007	0.008	-0.001	-0.18
truck	0.036	0.039	-0.003	-0.2
land ownership	0.657	0.646	0.011	0.252
animals ownership	0.394	0.382	0.012	0.281
workhead	0.748	0.807	-0.059	-1.566
workpouse	0.031	0.035	-0.004	-0.277
Household charac m	0.071	0.047	0.024	1.37
assets m	0.786	0.748	0.038	0.904
demographaphics m	0.049	0.043	0.006	0.386
N rooms	1.853	1.858	-0.005	-0.056
Hh age	40.047	41.449	-1.402	-1.174
Spouse age	32.659	33	-0.341	-0.301
N children 0-5	0.829	0.843	-0.014	-0.181
N children 6-12	1.697	1.752	-0.055	-0.571
N children 13-15	1.196	1.205	-0.009	-0.166
N children 16-20	0.757	0.685	0.072	1.054
N women 20-39	0.653	0.709	-0.056	-0.955
N women 40-59	0.392	0.382	0.01	0.256
N women 60	0.1	0.098	0.002	0.078
Continued on next page				

Table A.6 – continued from previous page

	Grade 9			z/t
	Treatment	Control	Diff	
N men 20-39	0.399	0.433	-0.034	-0.807
N men 40-59	0.537	0.567	-0.03	-0.734
N men 60	0.098	0.071	0.027	1.292

## 4.B Bootstrap procedure

Step 0:  $k=1$ ;

Step 1: For each sample of size  $n$ , take a random sample of  $n$  elements with replacement and estimate the logistic specification described in Section 4.3.3.

Step 2: for  $i=1$  to  $n$  compute the predicted values:

- $\beta'_k X_i^T$ , the index value if treated;
- $\beta'_k X_i^N$ , the index value if not treated;
- $p_i^T$  the probability of enrollment if treated;
- $p_i^N$  the probability of enrollment if not treated;

Define the coefficient vector associated with circumstances as  $\beta_c$ , the one associated with missing circumstances as  $\beta_m$  and the one associated with interactions between circumstances and treatment as  $\beta_{cm}$ .

Step 3: From step 2, some observations for which one or more coefficients  $\beta_k$  predict perfectly school enrollment were dropped, and therefore the corresponding predicted values  $\hat{p}_i^T$  and  $\hat{p}_i^N$  are not estimated. For these observations, we replace the estimated predicted values according to the following criteria:

- If  $\beta_c$  or  $\beta_m$  perfectly predicts being enrolled in school, replace  $p_i^T = p_i^N = 1$ . If  $\beta_c$  or  $\beta_m$  perfectly predicts not-being enrolled, replace  $p_i^T = p_i^N = 0$ .
- If  $\beta_{ct}$  perfectly predicts being enrolled in school, replace  $p_i^T = 1$ . If  $\beta_{ct}$  perfectly predicts not-being enrolled, replace  $p_i^N = 0$ .

Step 4: Use the values presented in step 3 to compute the following:

- (i) All the statistics mentioned in Tables 4.2 and 4.3 in the main text of the paper.
- (ii) The Lorenz and Generalized Lorenz coordinates at each decile for each grade

Store the results of (i) and (ii) in the  $k - th$  row of a matrix  $A$ .

Step 5: Repeat Steps 1-4 3000 times, and each time, store the results in the  $k - th$  row of matrix  $A$  so that we end up with matrix  $A$  of dimension  $[3000, B]$ , with  $B$  being the total number of statistics computed in each loop.

Step 6: Use the respective empirical distribution obtained for each statistic to construct the 99%, 95% and 90% confidence intervals for each of them.

Step 7: Finally, perform a test for Lorenz and Generalized Lorenz dominance on the basis of the vectors of ordinates. The test consist in posing the null as “nondominance” of the treated by the non-treated respective curve, such that, in case of rejection of the null, all that is left is dominance. To carry out the test, we compute the fraction ( $f$ ) of jointly positive values for the difference of the Lorenz and Generalized Lorenz coordinates, and interpret  $1 - f$  as the level of significance to reject the null.

## 4.C Generalized Lorenz

**Table A.7:** Generalized Lorenz coordinates per grade for treatment and controls

Grade in 1998	0	1	2	3	4	5	6	7	8	9
Decile 1 (T)	268.955	319.307	292.321	276.556	246.252	213.992	145.735	110.059	82.062	20.395
Decile 1 (NT)	268.147	300.111	279.923	255.830	209.624	224.687	98.247	110.520	74.343	14.162
Difference (T-N)	0.809	19.197	12.398	20.725	36.628	-10.696	47.488***	-0.461	7.719*	6.234*
Decile 2 (T)	572.950	650.106	602.475	564.211	500.877	434.210	308.548	222.386	167.658	44.763
Decile 2 (NT)	569.735	612.759	576.386	525.082	438.231	452.578	209.072	223.571	155.427	31.953
Difference (T-N)	3.215	37.347	26.089*	39.129***	62.646***	-18.368	99.476***	-1.185	12.231*	12.810*
Decile 3 (T)	898.905	985.904	917.421	856.441	759.435	666.126	485.372	338.022	255.626	71.330
Decile 3 (NT)	893.714	932.861	883.294	801.845	674.562	682.248	326.355	340.123	239.466	54.487
Difference (T-N)	5.191	53.043	34.127*	54.595***	84.873***	-16.123	159.017***	-2.101	16.160*	16.843*
Decile 4 (T)	1,244.863	1,322.397	1,235.532	1,150.133	1,020.969	903.354	667.020	455.793	345.082	99.697
Decile 4 (NT)	1,238.062	1,257.432	1,195.421	1,079.936	920.561	914.076	454.864	458.228	326.202	80.744
Difference (T-N)	6.801	64.965	40.111*	70.197***	100.408***	-10.722	212.156***	-2.435	18.880*	18.952*
Decile 5 (T)	1,597.925	1,661.827	1,561.993	1,446.290	1,284.304	1,144.231	863.277	574.346	435.030	130.075
Decile 5 (NT)	1,596.589	1,584.112	1,513.322	1,364.062	1,173.895	1,150.528	593.631	576.349	414.079	109.423
Difference (T-N)	1.336	77.715	48.671*	82.228***	110.409***	-6.297	269.647***	-2.002	20.951*	20.652*
Decile 6 (T)	1,967.781	2,002.885	1,891.052	1,747.058	1,550.066	1,386.048	1,072.427	695.176	525.995	162.425
Decile 6 (NT)	1,965.117	1,914.730	1,835.031	1,651.149	1,432.000	1,389.420	743.475	696.612	503.602	140.237
Difference (T-N)	2.664	88.155	56.021*	95.909***	118.066***	-3.372	328.951***	-1.435	22.393*	22.188
Decile 7 (T)	2,344.820	2,344.770	2,221.739	2,048.790	1,817.676	1,633.540	1,299.446	816.906	618.154	197.340
Decile 7 (NT)	2,340.224	2,247.375	2,159.389	1,942.944	1,694.257	1,632.185	917.015	817.978	594.660	176.493
Difference (T-N)	4.596	97.395	62.350*	105.846***	123.419***	1.355	382.431***	-1.072	23.494*	20.847
Decile 8 (T)	2,739.024	2,688.985	2,555.567	2,354.094	2,087.232	1,883.024	1,543.873	939.507	711.656	234.634
Decile 8 (NT)	2,724.944	2,584.907	2,487.511	2,239.610	1,962.440	1,878.256	1,121.268	940.219	687.255	215.676
Difference (T-N)	14.080	104.078	68.056*	114.484***	124.792***	4.768	422.605***	-0.712	24.401*	18.957
Decile 9 (T)	3,149.050	3,033.391	2,892.280	2,664.757	2,359.131	2,135.973	1,809.614	1,061.914	806.701	274.838
Decile 9 (NT)	3,130.184	2,926.530	2,823.956	2,545.436	2,236.558	2,128.589	1,351.678	1,062.361	781.225	257.085
Difference (T-N)	18.866	106.861	68.324*	119.321***	122.573***	7.384	457.936***	-0.447	25.476*	17.753
Decile 10 (T)	3,579.128	3,378.930	3,232.098	2,979.497	2,635.260	2,392.714	2,114.512	1,186.323	902.830	321.432
Decile 10 (NT)	3,558.825	3,271.238	3,165.490	2,858.615	2,517.366	2,384.422	1,625.953	1,187.184	876.949	306.902
Difference (T-N)	20.303	107.692	66.608*	120.882***	117.894***	8.292	488.559***	-0.861	25.880*	14.530
Positive treatment effect (joint test)				**	***		***			

*Notes: The stars behind the coefficients indicate the difference between treated and non-treated is statistically significant. \*\*\* (\*\*) [\*] means significant at 1 (5) [10] percent. Level of significance based on Bootstrap confidence intervals. (T) stands for treated while (NT) for non-treated. A joint test for dominance at the end of the table indicates if the difference between treated and non-treated is jointly positive for all deciles. Source: Authors' calculations.*

## 4.D Lorenz

Table A.8: Lorenz coordinates per grade for treatment and controls

Grade in 1998	0	1	2	3	4	5	6	7	8	9
Decile 1 (T)	0.0751	0.0945	0.0904	0.0928	0.0934	0.0894	0.0689	0.0928	0.0917	0.0635
Decile 1 (NT)	0.0753	0.0917	0.0884	0.0895	0.0833	0.0942	0.0604	0.0932	0.0856	0.0462
Difference (T-N)	-0.0002	0.0028	0.0020	0.0033	0.0102***	-0.0048	0.0085**	-0.0004	0.0062*	0.0173*
Decile 2 (T)	0.1601	0.1924	0.1864	0.1894	0.1901	0.1815	0.1459	0.1875	0.1874	0.1395
Decile 2 (NT)	0.1601	0.1873	0.1821	0.1837	0.1741	0.1898	0.1286	0.1885	0.1789	0.1043
Difference (T-N)	0.0000	0.0051	0.0043	0.0057**	0.0160***	-0.0083	0.0173**	-0.0010	0.0085*	0.0352**
Decile 3 (T)	0.2512	0.2918	0.2838	0.2874	0.2882	0.2784	0.2295	0.2849	0.2857	0.2222
Decile 3 (NT)	0.2511	0.2852	0.2790	0.2805	0.2680	0.2861	0.2007	0.2867	0.2756	0.1778
Difference (T-N)	0.0000	0.0066	0.0048	0.0069**	0.0202***	-0.0077	0.0288***	-0.0018	0.0102*	0.0444**
Decile 4 (T)	0.3478	0.3914	0.3823	0.3860	0.3874	0.3775	0.3154	0.3842	0.3857	0.3106
Decile 4 (NT)	0.3479	0.3844	0.3776	0.3778	0.3657	0.3834	0.2798	0.3863	0.3754	0.2635
Difference (T-N)	-0.0001	0.0070	0.0046	0.0082**	0.0217***	-0.0058	0.0357***	-0.0021	0.0103*	0.0471**
Decile 5 (T)	0.4465	0.4918	0.4833	0.4854	0.4874	0.4782	0.4083	0.4841	0.4863	0.4053
Decile 5 (NT)	0.4486	0.4843	0.4781	0.4772	0.4663	0.4825	0.3651	0.4858	0.4765	0.3570
Difference (T-N)	-0.0022	0.0076	0.0052*	0.0082**	0.0210***	-0.0043	0.0432***	-0.0017	0.0098*	0.0482*
Decile 6 (T)	0.5498	0.5928	0.5851	0.5864	0.5882	0.5793	0.5072	0.5860	0.5880	0.5060
Decile 6 (NT)	0.5522	0.5853	0.5797	0.5776	0.5688	0.5827	0.4573	0.5872	0.5795	0.4576
Difference (T-N)	-0.0024	0.0074	0.0054*	0.0088**	0.0194***	-0.0034	0.0499***	-0.0012	0.0084*	0.0484*
Decile 7 (T)	0.6551	0.6939	0.6874	0.6876	0.6898	0.6827	0.6145	0.6886	0.6910	0.6148
Decile 7 (NT)	0.6539	0.6870	0.6822	0.6797	0.6730	0.6845	0.5640	0.6895	0.6843	0.5759
Difference (T-N)	0.0013	0.0069	0.0052**	0.0079**	0.0167***	-0.0018	0.0506***	-0.0009	0.0066*	0.0389*
Decile 8 (T)	0.7653	0.7958	0.7907	0.7901	0.7920	0.7870	0.7301	0.7919	0.7955	0.7310
Decile 8 (NT)	0.7657	0.7902	0.7858	0.7835	0.7796	0.7877	0.6896	0.7925	0.7909	0.7038
Difference (T-N)	-0.0004	0.0056	0.0049**	0.0066**	0.0125***	-0.0007	0.0405***	-0.0006	0.0046*	0.0272*
Decile 9 (T)	0.8798	0.8977	0.8949	0.8944	0.8952	0.8927	0.8558	0.8951	0.9017	0.8563
Decile 9 (NT)	0.8796	0.8946	0.8921	0.8904	0.8885	0.8927	0.8313	0.8955	0.8990	0.8389
Difference (T-N)	0.0003	0.0031	0.0028**	0.0039***	0.0068***	0.0000	0.0245***	-0.0004	0.0027*	0.0174*
Positive treatment effect (joint test)				*	***		**			
(joint test)										

Notes: The stars behind the coefficients indicate the difference between treated and non-treated is statistically significant. \*\*\* (\*\*) [\*] means significant at 1 (5) [10] percent. Level of significance based on Bootstrap confidence intervals. (T) stands for treated while (NT) for non-treated. A joint test for dominance at the end of the table indicates if the difference between treated and non-treated is jointly positive for all deciles. Source: Authors' calculations.